1st Class@Number 1

Effectiveness trial

Evaluation Report

June 2025

Aarti Sahasranaman, Chris Morton, Eleanor Bradley, Emma Moore, Pippa Lord

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Contents

About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research (NFER). The trial director and principal investigator for this study was Pippa Lord, Research Director and Early Years Lead. Aarti Sahasranaman, Senior Research Manager, led the evaluation team and the impact evaluation. Emma Moore, Senior Research Manager, led the implementation and process evaluation (IPE). They were supported by Chris Morton as trial statistician, Eleanor Bradley as IPE researcher, and Emily Jones as advisor on assessments. Kathryn Hurd led the research operations with Jishi Jose, Lydia Wallis, and Rob Green as operations researchers.

Contact details

National Foundation for Educational Research
The Mere,
Upton Park,
Slough, Berkshire SL1 2DQ,
United Kingdom
Telephone: +44 (0)1753 574123
Email: a.sahasranaman@nfer.ac.uk

Acknowledgements

We are grateful to all the primary schools that participated in this evaluation. We are especially thankful to the school staff members who completed the baseline assessments and coordinated the endpoint assessments with NFER test administrators and the staff and pupils who took part in the case study visits and interviews.

We would like to give a special mention to the staff at Edge Hill University who have worked with us to support the evaluation whenever needed, namely Yvonne Panteli, Peter Hick, Jane Moore, Lisa Leach, Linda Lavagna-Slater, and Jonathan Glazzard.

We would like to thank the team at the Education Endowment Foundation, including Ignacia Valenzuela, Ben Simuyandi, Beverley Jennings, and Celeste Cheung.

ONS SRS publication status and statistical results

Executive summary

The project

1stClass@Number 1 is a small group intervention developed by the Every Child Counts (ECC) team at Edge Hill University. It is designed for children who are experiencing moderate difficulties in maths requiring further support at the level of the Year 1 curriculum. The intervention is designed to be delivered by trained teaching assistants (TAs) to groups of four Year 2 pupils in 30 sessions (six sessions for each of the five topics) over a ten-week period. TAs receive six training sessions (one full day covering two sessions and four half-days covering one session each) and start delivering topics to pupils after receiving training for that topic. TAs are provided with detailed session guidance and extensive resources to support delivery of the topics to pupils. Schools also nominate a Link Teacher who receives training to support the TA with implementation and provide feedback.

A previous effectiveness trial commissioned by the Education Endowment Foundation (EEF) in 2018 found evidence of positive impact of the intervention on pupils' attainment. Although this finding had a high security rating, it was not statistically significant, likely due to inadequate sample size. So, the evaluator could not rule out the intervention not having an effect. Furthermore, the trial did not find any evidence that the intervention improved attainment outcomes for FSM-eligible pupils (those in receipt of free school meals).

The study was funded by the EEF through the Department for Education's Accelerator Fund. The EEF has commissioned this effectiveness trial to assess the impact of 1stClass@Number 1, with a particular focus on FSM-eligible pupils. The trial is powered to detect an impact for all pupils (primary research question) and for FSM-eligible pupils. 1stClass@Number 1 was evaluated in this trial using a within-school pupil-randomised design (that is, it was a multisite trial). Two hundred and twenty-six schools were recruited and eight selected pupils from within each school were randomised on a 1:1 basis into the intervention and control arms. In response to the challenge posed by recruiting this large number of schools, the trial was run in the same academic year (2023/2024) with the 226 schools split into two cohorts. Recruitment commenced in February 2023 and was completed for both cohorts in October 2023. Training and delivery started in November 2023 for Cohort 1 and in January 2024 for Cohort 2. The primary outcome measure used was the Quantitative Reasoning Test (QRT: Nunes et al., 2015). Endpoint assessments were completed by NFER test administrators in March 2024 for Cohort 1 and in June 2024 for Cohort 2.

Table 1: Key conclusions

Key conclusions	
1.	Pupils receiving 1stClass@Number 1 made the equivalent of two additional months' progress, on average, compared to pupils who did not receive it. This result has a high security rating.
2.	Pupils eligible for free school meals (FSM) receiving 1stClass@Number 1 made the equivalent of two additional months' progress, on average, compared to FSM-eligible pupils who did not receive it.
3.	TAs perceived the training content, delivery, and materials to be of high quality and reported that they enjoyed delivering the intervention to pupils. In over 80% of schools, the same TA attended all of the delivery-training sessions, that is, there was continuity of TA, which is important for effective delivery.
4.	Pupils reported enjoying the intervention activities and interactive games. TAs and Link Teachers felt pupils were engaged and that the programme had a positive impact on pupils' maths attainment. Just over 60% of pupils attended at least five of the six sessions for each topic—a threshold which was felt to be important for the success of the programme.
5.	The key features of 1stClass@Number 1 that are likely to have led to the positive impact on pupils' attainment include its modular nature with each topic building on previous ones, its manualised nature, high quality training, high quality small group tutoring with optimum group size and session frequency, and the ability to be adapted to pupils' needs.

EEF security rating

These findings have a high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed, well-powered, two-arm within school randomised controlled trial where 13.2% of the pupils who started the trial were not included the final analysis: in the majority of cases this was due to the pupil either leaving the school or being absent on the day of the test. The

pupils in the 1stClass@Number 1 intervention group were similar to those in the comparison group in terms of prior attainment.

Additional findings

Pupils who received 1stClass@Number 1 made, on average, the equivalent of two additional months of progress compared to pupils who did not receive the programme. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes positive effects of one to three months of additional progress. The data suggests it is highly likely that the intervention improves maths attainment as measured by the QRT. In contrast to findings from the previous trial, this trial—which was well-powered for FSM pupils—found that FSM pupils who received 1stClass@Number 1 made the equivalent of two additional months of progress compared to those in the control group.

There is considerable evidence from the implementation and process evaluation (IPE) that the short- and intermediate-term outcomes for TAs, Link Teachers, and pupils set out in the intervention logic model have been achieved. For example, TAs, Link Teachers, and pupils reported that pupils' maths skills and confidence had increased. TAs reported improved knowledge and confidence teaching mathematics and a greater awareness of correct mathematical language. Preparation time for TAs and alignment of sessions to the national curriculum were identified as important moderators.

Given that pupils received the programme in addition to their usual maths lessons and in small groups, it is reasonable to question whether the observed effects are solely due to pupils experiencing more maths and/or to high-quality small group tuition. Pupils in the intervention group did indeed receive more maths instruction, and the IPE confirms that 1stClass@Number 1 includes many of the key features of small group interventions known to be effective including optimal group size (in this case, four pupils), and frequent sessions (in this case, three 30-minute sessions per week) delivered over a ten- to twelve-week period. However, the IPE also highlights unique features of 1stClass@Number 1 that might have contributed to its impact, including the gradual building of knowledge where each topic builds on previous ones, high quality training, resources, and support, and the provision of delivery adaptations to tailor to pupils' needs.

The use of TAs to deliver a highly manualised intervention is an efficient and potentially cost-effective use of limited resources in schools. By providing high quality training supported by comprehensive lesson plans, scripted questions, and resources, 1stClass@Number 1 allows TAs to independently deliver small group tutoring thereby allowing teachers to focus on classroom teaching. The intervention builds capacity within schools for TAs to support pupils facing moderate difficulties, reducing the need for potentially more expensive provision by external tutors. Most schools reported a continuing need for maths support, and given their positive experiences of the intervention intend to continue using the programme beyond the trial—both with Year 2 pupils in the control group and with next year's Year 2 pupils.

Cost

The average cost of implementing 1stClass@Number 1 for one school is around £745 per year or £93 per pupil per year when averaged over three years (on the assumption that the intervention is delivered to two groups of four pupils each year). There is a reduction in costs in years two and three compared to the first year because training does not have to be delivered again and some of the materials and resources can continue to be used.

Impact

Table 2: Summary of impact on primary outcomes

Outcome/group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of Pupils (intervention; control)	p-value	EEF cost rating
Maths attainment (all pupils)	0.12 (0.05, 0.19)	2	🔒🔒🔒🔒🔓	1560 (778; 782)	< 0.001	**£ £** £ £ £
Maths attainment (FSM pupils)	0.11 (0.02, 0.20)	2	N/A	993 (494; 499)	0.015	**£ £** £ £ £

Introduction

Background

Mathematical attainment is a determinant of educational progress, socioeconomic status, employability, physical and mental health, and financial stability (Parsons and Bynner, 2005; Duncan et al., 2007; Gerardi, Goette and Meier, 2013; Ritchie and Bates, 2013). Mathematical attainment in primary school is crucial because the links between Key Stage 2 performance and later educational achievement are particularly strong (Menzies, Ramaiah and Boulton, 2021). This means that pupils who perform poorly in maths in primary school are unlikely to turn this around and perform better at secondary school. Poorer education outcomes may eventually have a bearing on wider outcomes such as economic outcomes, health outcomes, and life chances. In 2023, 70% of all pupils (and 75% of non-disadvantaged pupils) in Key Stage 1 met the expected standard in maths (GOV.UK, 2023). In comparison, only 56% of pupils from disadvantaged backgrounds met the expected standard, highlighting the attainment gap between disadvantaged pupils and their more affluent peers. These attainment gaps appear early and often persist through the various stages of schooling (Farquharson, McNally and Tahir, 2022).

One to one tutoring is an effective strategy to improve pupil outcomes, with an average impact of five additional months' progress across different school phases and subjects (Harrison and Higgins, 2023). However, delivering one to one tuition is expensive, and this is especially true when tutoring is delivered by qualified teachers. Small group tuition, in comparison, has a slightly smaller impact, with an average of an additional four months' progress (Harrison and Higgins, 2023). Impact tends to be higher for primary schools (four months' additional progress) compared to secondary schools (two months' additional progress), which has fewer studies overall. Most studies of small group tuition focus on reading for which there is a greater average impact (four additional months' progress) compared to maths (three months). However, small group tuition may be a cost-effective solution for schools especially when tutoring is delivered by trained teaching assistants (TAs). When TAs deliver small group tutoring in structured settings with high quality training and support, this results in consistent positive impacts on pupils' attainment (Sharples, Webster and Blatchford, 2021). 1stClass@Number 1 is one of three mathematical interventions developed by Every Child Counts (ECC),[1] which was set up with support from the Department for Education (DfE) in 2008 and run on a not-for-profit basis by Edge Hill University (EHU). The 1stClass@Number interventions have been widely used by schools and, so far, have supported over 55,000 pupils in Years 1 to 11 across 4,000 schools.[2]

1stClass@Number 1 is a small group intervention for children who need further support at the level of the Year 1 curriculum. It is delivered in 30 sessions by trained TAs to groups of four Year 2 pupils. TAs receive six training sessions from EHU's ECC trainers (see Intervention section for more detail on the topics covered in each training session). TAs start delivering topics to pupils after receiving training for that topic. Training sessions are scheduled every two to four weeks to allow TAs sufficient time to deliver the topic to pupils. TAs are provided with detailed session guidance and extensive resources to support delivery of the topics to pupils.

EHU runs its own internal evaluation of the 1stClass@Number interventions wherein schools are encouraged to submit the pre- and post-test outcome scores of pupils to EHU's online system, which, in turn, provides a detailed analysis of the scores. [3] EHU recommends that schools delivering the 1stClass@Number interventions administer the standardised Sandwell Early Numeracy Test, Revised (SENT-R) at baseline and endpoint. Although not a rigorous randomised controlled trial (RCT), this internal evaluation showed that pupils made, on average, a number age gain of 13 months in only four months.[4] Ninety-three percent of pupils who received the intervention showed more confidence and interest in learning mathematics in class.

[1] About ECC - Every Child Counts (edgehill.ac.uk) https://ehu.ac.uk/ecc

[2] 1stClass@Number - Every Child Counts (edgehill.ac.uk) https://ehu.ac.uk/1stclassnumber

[3] 1stClass@Number - Every Child Counts (edgehill.ac.uk) https://ehu.ac.uk/1stclassnumber

[4] Note that the internal evaluation by EHU included pupils from Years 1 to 11 who received the 1stClass@Number interventions in previous years. It did not include pupils from this evaluation.

Given the positive impact demonstrated by EHU's internal evaluation and the widespread use of the 1stClass@Number 1 interventions in schools, the Education Endowment Foundation (the EEF) commissioned the evaluation of 1stClass@Number 1 through an effectiveness trial (Nunes et al., 2018). This school-randomised RCT with 130 schools in the 2017/2018 academic year found that pupils who received 1stClass@Number 1 made, on average, two additional months' progress in maths compared to pupils who did not receive the intervention. Although this finding had a high security rating, it was not statistically significant, likely due to inadequate sample size. There was also no evidence to suggest that this progress in maths translated to improved KS1 maths outcomes. Among FSM pupils,[5] those who received the intervention did not make any additional progress in maths compared to those who did not receive the intervention; in fact, there was indicative evidence to suggest that 1stClass@Number 1 was not as effective for FSM pupils as for pupils not eligible for FSM.

Accessing the DfE's Accelerator Fund developed to close the disadvantage gap, the EEF commissioned a formative evaluation by the Behavioural Insights Team (BIT) in 2022 to further explore this differential impact of 1stClass@Number 1 and to better understand the pupil selection mechanisms employed by schools.[6] The evaluators analysed pupils' pre- and post- test scores on the SENT-R collected by EHU and found no clear evidence that 1stClass@Number 1 was less beneficial to FSM pupils. The evaluation also found that schools selected a 'not unsubstantial' proportion of pupils who were working above the recommended level in maths for the intervention (that is, who did not appear to need additional support at the level of the Year 1 curriculum in terms of their age-related progress in maths, although we note that in a pandemic context, schools may also have been selecting pupils using other criteria). The BIT evaluators, therefore, recommended that schools be provided with clear guidance on selecting pupils to ensure the selection of pupils who were most likely to benefit from the programme.

In light of evidence from the previous effectiveness trial and the findings of the formative evaluation conducted by BIT, the EEF commissioned this effectiveness trial to rigorously assess the impact of 1stClass@Number 1 among Year 2 pupils, with a particular focus on FSM-eligible pupils. Although our primary research question is to assess the impact of 1stClass@Number 1 on all pupils (in line with the EEF's remit across all trials), this trial is designed to be powered for the FSM subgroup, allowing us to assess the intervention's impact on this specific subgroup. In response to BIT's recommendation that schools be provided with clear guidance on selecting pupils, this trial has also implemented a more objective selection process (detailed in the Methods section). Finally, one of the shortcomings of the earlier effectiveness trial was that it was underpowered and we note that the findings were not statistically significant. The power calculations for that trial assumed a pre test/post test correlation of 0.7 whereas the correlation for nominated pupils in that trial was found to be more modest at 0.29. This lower correlation was not unexpected as the QRT was used both to select pupils into the trial (in combination with other eligibility criteria) and as a baseline measure, thereby curtailing the range of pre-test QRT scores. Assuming a pre test/post test correlation of 0.29, our power calculations precluded randomisation at the school level as this would require a very large number of schools to be recruited (over 500) rendering the trial practically and financially unfeasible. We have, therefore, designed this trial with pupil-level randomisation within schools—possible due to the nature of the small-group extraction of pupils from class to receive the intervention in specified sessions. The pupil-randomised design included further strategies for avoiding contamination between intervention and control group pupils within schools (see Recruitment and Pupil Data Collection below).

Our implementation and process evaluation complements the impact evaluation and focuses on areas such as the pupil selection guidance, training model, and the pupil randomised design. It also explores the key mediators and moderators to understand the mechanisms by which the intervention will produce outcomes. Given the focus on FSM pupils in the impact evaluation, the IPE includes some lines of enquiry to explore any perceived differences for disadvantaged pupils (for example, in terms of implementation, engagement, and perceived outcomes) although we note that the intervention itself is for any pupil below expected standards (that is, it is not an intervention specifically targeted at FSM pupils).

[5] Note that pupils are considered FSM-eligible only if an active claim for FSM eligibility has been made by their parents/carers.

[6] The findings from this study were reported in an internal presentation to the EEF that was shared with potential evaluators at the time of submission of the expression of interest to deliver this evaluation.

Intervention

1st Class@Number 1 is a maths intervention for Year 2 pupils developed by EHU. It comprises six training sessions for TAs in which EHU's ECC trainers share the programme content, discuss common misconceptions, and reflect on implementation and progress. The first and last of these training sessions are also attended by a designated 'Link Teacher' from each school whose role is to facilitate the running of the intervention in their school. The Link Teacher is typically a senior member of staff such as the school's maths lead, a head of key stage, assistant headteacher, or potentially the headteacher. TAs then deliver a total of 30 intervention sessions divided into five topics to a total of four pupils at their school. Details of the programme are provided using the Template for Intervention Description and Replication (TIDieR) checklist below.

Recruitment for the trial began in February 2023 and was undertaken by EHU and was initially due to finish in July 2023 to allow the delivery phase of the trial to begin in September 2023. However, challenges with recruiting the large number of schools required to sufficiently power the trial for FSM pupils led to the decision to switch to a two-cohort design. This meant that recruitment for the second cohort of schools could continue until October 2023 and the delivery phase for Cohort 2 was scheduled to begin in January 2024. In addition, the decision was made to include one online training group in Cohort 2 to allow the school recruitment target to be met.

TIDieR checklist

Intervention name

1st Class@Number 1.

Why? —rationale/theory for the programme

Small group maths tuition has previously been suggested to be effective for improving pupils' attainment, especially when it is targeted at pupils' needs. EHU developed 1st Class@Number 1 to help Year 2 pupils with moderate difficulties—that is, those who are working approximately one year behind age-related-expectations (ARE) in maths—to catch-up with their peers. The intervention is designed to be delivered by TAs. It is based on connected learning experiences and the idea that children develop their mathematical understanding through a combination of concrete experiences, language, pictures, and symbols (for example, those used for numbers, operations, and equality).

The programme comprises 30 intervention sessions for pupils which are divided into five topic areas:

- 'all about number';
- 'exploring place value';
- addition and subtraction (1);
- addition and subtraction (2); and
- 'towards multiplication and division'.

These topic areas align with the national curriculum and are designed to address the most common errors and misconceptions in primary maths learning. Each topic comprises six sessions designed to build upon one another and allow TAs to assimilate pupils' understanding of each topic area prior to delivering the rest of the topic content.

Who?—recipients

1st Class@Number 1 is targeted at Year 2 pupils (aged six to seven) with moderate difficulties in mathematics to enable them to catch up with their peers. Pupils are eligible for the programme if they are working at a level approximately one year below ARE at the end of Year 1. For example, pupils may:

- be able to count forwards in ones to ten;

- have some knowledge of number facts and have some understanding of the composition of number within ten;
- be able to perform simple addition and subtraction calculations using counting all and counting out approaches; and
- be able to read and write numbers but lack secure understanding of their magnitude and quantity.

However, they may struggle with mathematical vocabulary, lack confidence in mathematics, and be reluctant to talk about their mathematical learning.

Pupils were screened for trial eligibility using the SENT-R. Details of how pupils were selected to the trial based on their SENT-R scores are provided in the Participant Selection section. Pupils who were selected for 1st Class@Number 1 could not participate in any other maths interventions during the trial. In addition, since this trial was funded by the DfE Accelerator Fund, 50% of schools had to be located in Education Investment Areas (EIAs). In addition, because of the trial's focus on disadvantaged pupils—those eligible for free school meals—all schools were encouraged, amongst their nominations for screening, to nominate FSM pupils who met the criteria described above.[7]

EHU (the delivery partner) was responsible for recruiting schools to the trial. To be eligible to participate in the trial schools had to:

- be an infant or primary school with children in Year 2 as of 1 September 2023;
- *not* be taking part in either the ARK Mathematics Mastery or Mathematical Reasoning programme, which were also funded by the EEF in 2023/2024;
- *not* currently have a TA working in the school who had already completed any 1st Class@Number 1 training;
- be able to nominate a TA and Link Teacher who could attend face to face training in one of the 16 training locations; and
- be able to pay the subsidised training cost of £200 (usual cost is £1,100).

What?—materials/resources

At the first training session TAs received a set of handbooks containing an overview of the programme and guide to organising their time during intervention delivery, detailed lesson plans for each topic, as well as a resource box containing almost all of the resources needed to deliver the programme. The resources also contained 'special delivery' activities that pupils could take home to do with their parents/carers. The only materials not included were those readily available in schools (such as coins and linking cubes). Schools were also given access to an online portal where they could download additional copies of the consumable materials to give out to pupils.

What?—processes, activities

TAs received six training sessions that were delivered by EHU's ECC trainers accredited by EHU. The training sessions took place roughly every two to four weeks. The first two sessions were delivered in a full day and the remaining four each lasted a half-day. The first five training sessions ran alongside the intervention delivery, which meant that TAs were trained on the intervention topic by topic. These sessions focused on how the intervention works, how to deliver it to pupils including making adaptations to meet pupils' needs and developing TAs' own understanding of mathematics. The final training session focused on reflection and sharing of practice.

Link Teachers—whose role was to support TAs and the strategic implementation of the programme—received two half-day training sessions that coincided with the first and last training sessions for TAs. These sessions focused on

[7] For all 'pre-analysis' purposes (including recruitment, pupil selection, and stratified randomisation) the FSM variable was created by asking schools which of their pupils were currently eligible for FSM.

managing and supporting the implementation of 1stClass@Number 1 and sharing learning from the programme with colleagues after completion of programme delivery.

Training model

All training sessions were delivered face to face except for a single online training group in Cohort 2 (see below). Each session was followed by a break to allow for intervention delivery, following which training was provided in the next topic. In addition to preparation for the next topic, each TA training session also included a feedback and review component where TAs could discuss their experience of delivering the previous topic to pupils and share any lessons learned.

The training sessions were structured as follows:

- the first training day consisted of two half-day sessions: session one was for both TAs and Link Teachers and provided an introduction to 1stClass@Number 1; session two was for TAs only and focused on preparation for topic one, 'all about numbers';
- the remaining training days—two to five—consisted of four further half-day training sessions (sessions three to six):
    - session three prepared TAs to deliver topic two: 'place value';
    - session four prepared TAs to deliver topics three and four: addition and subtraction (1) and (2); and
    - and Session 5 prepared TAs to deliver topic five: multiplication and division; and
    - session six consisted of one half-day session for both TAs and Link Teachers where trainers discussed how to share the experiences of participating in the intervention at staff meetings in schools and how TAs and Link Teachers could consider offering 1stClass@Number 1 for future Year 2 cohorts.

A small online training group was recruited to support recruitment targets for Cohort 2 and had the benefit of catering to schools that were too far from one of the in-person venues. The structure and content of the online training model was identical to that of the face to face training model, meaning that the key difference was that the mode of delivery was online. All group activities during the training sessions were designed to take place in virtual break-out rooms where smaller groups of trainees could interact with each other. It is important to note that EHU had already developed and used this online training model as part of its routine delivery of 1stClass@Number 1, and it was agreed to instigate this to support recruitment targets.

Attendance at training sessions is routinely monitored by EHU. For the purpose of the trial, trainers endeavoured to offer a catch-up session (either face to face or remotely) to TAs or Link Teachers who missed a training session. These typically took place in the week following scheduled training delivery. If TAs or Link Teachers were also unable to attend this catch-up session, then they could contact trainers to address any queries and TAs could use the topic handbook to deliver the intervention sessions.

All schools also received one individual visit by their trainer which took place before the midpoint of the intervention to offer support and help quality assure delivery. Training and support visits were provided by EHU's ECC 1stClass@Number trainers, experienced former teachers with expertise in the delivery of maths professional development to school colleagues. Trainers were also available for TAs or Link Teachers to contact with any queries by telephone or email throughout the intervention period. Each group of ten to fifteen TAs, who would come together for each training session, had a trainer who was responsible for them. The total number of training groups based on the number of schools required for the trial was 17 (and these groups were delivered by 12 trainers).

Who? —implementers

1stClass@Number 1 is one of three mathematics interventions developed and delivered by EHU. EHU's ECC trainers delivered the training to TAs and Link Teachers. TAs delivered the intervention to pupils. Link Teachers were expected to meet with TAs once a week to review progress, plan upcoming sessions, and provide feedback.

Trainers

All trainers were part of ECC and accredited by EHU. EHU's ECC helps children in need of additional support through a suite of maths interventions. All trainers had to have previously delivered 1stClass@Number 1 themselves and completed the university accreditation process in order to be eligible to become a trainer and many of the trainers were independent mathematics consultants.

TAs

Schools were asked to nominate a TA to attend the training sessions and deliver the intervention to pupils. Guidance provided in the 1stClass@Number 1 handbook states that a suitable TA should have prior experience supporting pupils' mathematical development in the appropriate age phase (Key Stage 1), be able to make decisions when planning and teaching pupils based on an understanding of their needs with the support of a Link Teacher and be able to attend all the training sessions.

Link Teachers

Schools were also asked to nominate a 'Link Teacher' to support the TA delivering the programme. The intervention handbook recommends that the Link Teacher is a member of staff with suitable experience to support with mathematics, is someone able to provide strategic leadership for the programme and ensure that the intervention is effectively managed/implemented to maximise impact, and someone who has time to support the TA and liaise with class teachers and senior managers, as well as attend the two training sessions.

How? —mode of delivery

1stClass@Number 1 intervention sessions are designed to be delivered face to face to groups of four pupils during the school day.

The training sessions for TAs and Link Teachers took place in-person in a central location within each training region. The exception to this was the online group where training was delivered fully online. Additional catch-up sessions were provided online for TAs or Link Teachers who were unable to attend the original training session.

Where? —setting of the intervention

Ideally the interventions sessions were meant to be delivered in a separate, designated space within participating schools where the group could sit together around a table. This was to prevent distractions and ensure the noise made during the games did not distract other pupils. It also meant that the pupils' work could be displayed on the walls (note strategies for minimising contamination in this pupil-randomised trial included requesting that such displays were not visible outside of sessions).

For the trial, it was originally anticipated that intervention schools would be located in 16 regions across England: Birmingham, North Manchester, South Manchester, Liverpool, Lancashire, Rotherham, Cambridgeshire, East Sussex, North East England, Peterborough, Wiltshire/Berkshire, Derbyshire/Nottinghamshire, North Somerset/South Gloucestershire, Dorset, and the Isle of Wight and a second training group covering different areas in Liverpool/Lancashire. These training regions were chosen based on the availability of EHU's ECC trainers. In addition, some regions are also EIAs, regions of focus for recruitment in this trial.

However, four of these groups—located in Peterborough, Wiltshire/Berkshire, the Isle of Wight, and North Somerset/South Gloucestershire—were cancelled due to low demand and numbers were made up from other regions where demand was higher. A total of five schools from these four regions were recruited (that is, signed memorandum of understanding) and one chose to continue with the trial as part of the online group. The other four schools chose to withdraw during the recruitment phase. In addition, a small number of schools were recruited from outside of these regions and made up the online training group. As noted above, the content of these training sessions was the same, but TAs and Link Teachers attended the training online. The recruitment of schools to this online training group was necessary for meeting the target recruitment numbers for the trial to ensure sufficient statistical power.

The inclusion of training groups in Cohort 1 or Cohort 2 was determined by how quickly each training group was filled. Groups that were filled quicker formed Cohort 1 whereas groups where recruitment was slower formed Cohort 2. This

meant that Cohort 1 consisted of training groups in the following regions: Birmingham, North Manchester, South Manchester, Liverpool, Lancashire, Rotherham, Cambridgeshire, East Sussex, Derbyshire/Nottinghamshire, West Sussex with a second training group in Liverpool/Lancashire. Cohort 2 consisted of training groups in North East England, North Manchester (second group), Birmingham (second group), Derby, Dorset, and the online training group.

When and how much? —dosage and duration

The intervention consisted of 30 sessions designed to be delivered three times per week over ten weeks. Each session was designed to last approximately 30 minutes. Sessions were meant to be delivered during normal school hours in addition to regular mathematics lessons. This meant that the pupils would spend approximately 15 additional hours on mathematics over the course of the programme. Note that training days were interspersed between the topics so the total delivery time for the whole intervention including training was approximately 15 weeks.

Each of the five topics within the intervention contains six sessions designed to be delivered over a two-week period. The structure of the sessions within each topic is the same: each starts with an introductory 'setting the scene' activity to allow TAs to understand pupils' current knowledge in that topic area before the main content of that topic is delivered.

The intervention handbook provides information about allowing TAs sufficient time to prepare for the sessions and how to timetable the sessions for pupils. If specific pupils within the group are struggling, then there is the option for TAs to deliver additional intervention sessions to these pupils to help them keep up with the group. TAs were asked to log pupils' attendance at each intervention session and share this data with the research team.

During the trial the intervention was delivered by TAs in schools between November 2023 and May 2024. Cohort 1 schools delivered the intervention from November 2023 to March 2024; Cohort 2 from January 2024 to May 2024.

Adaptations

Each topic begins with simple diagnostic tasks to help TAs understand what pupils already know and any potential gaps in their knowledge so they can tailor the subsequent topic sessions to pupils' needs. All the children still complete the same activities but adaptations to make the activities easier or more difficult (for example, by using smaller or larger numbers) are included in the topic handbook.

Strategies to support successful implementation

TAs were provided with thorough training on the intervention as well as almost all of the resources (see TiDieR—What? for details) they need to deliver the intervention including detailed lesson plans for each session.

Schools selected Link Teachers to support the TAs to deliver the intervention and quality assure their delivery.

All schools that took part in the trial received a visit from their trainer either online or in-person during the early stages of delivery. This was to offer additional support and for trainers to check-in on how things were going and answer any questions. While this support was provided to all schools during the trial, outside of the trial this visit is an optional extra for schools.

Parents/carers were provided with information about the intervention and pupils were given 'special delivery' activities which they could take home to do with their parents/carers.

Logic model

Figure 1 shows the proposed logic model for 1st Class@Number 1. It outlines the target population—Year 2 pupils struggling with mathematics—and the activities, outputs, short-term, and intermediate outcomes that lead to the outcome of improved maths attainment both immediately after the intervention and in the longer term. The intervention involves a training programme for TAs and Link Teachers as well as intervention delivery to pupils, so each group of participants involved in the intervention activities is included as a separate row within the logic model. This categorisation allows the representation of the different moderators that are applicable to each participant group and could have an impact on overall attainment. Both short-term and intermediate outcomes are included in the logic model to denote the ordering in which outcomes are posited to occur. These have not all been widely explored in previous research so are

a key area of interest for this trial. The IPE therefore explored these short-term and intermediate outcomes in more detail alongside the moderators that may lead to differential outcomes for FSM pupils/all pupils.

1stClass@Number 1 is mapped to the school curriculum and is designed to address the most common maths difficulties that pupils experience at this stage of their education. A key assumption of the logic model, therefore, is that the right pupils are selected for the programme. In response to BIT's formative study, guidance on pupil selection was provided to schools. For the purposes of the trial, the SENT-R screening test (previously recommended in EHU's guidance) was made mandatory, thereby making the pupil selection process more objective.

TA selection is also an important moderator, represented in the TIDIeR checklist in terms of TA's prior experience in supporting maths and their ability to engage with the maths curriculum and subject knowledge in the training sessions themselves. It is expected that their subject and pedagogical knowledge of maths will improve during the intervention, and that they will be confident to deliver 1stClass@Number 1 and tailor sessions to their pupils (short-term outcomes). As set out in the logic model, it is then expected that this will lead to increased confidence to support pupils in maths more widely (intermediate outcomes).

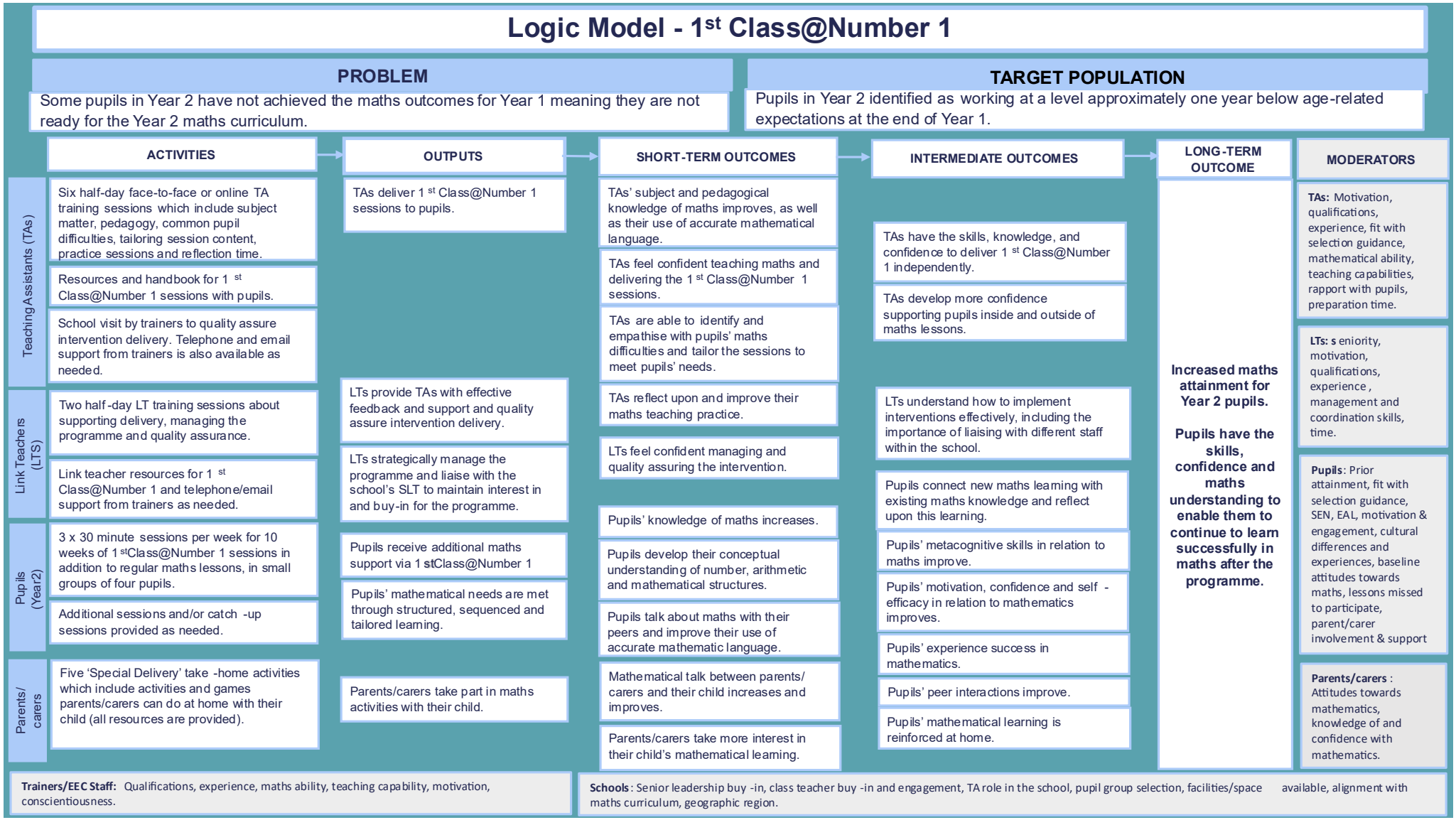
The training TAs received took place alongside the intervention delivery. Therefore, the logic model makes the assumptions that:

- the training delivery is consistent and high-quality;
- TAs will attend all the training sessions; and
- the time they have between the training session and delivering the relevant sessions for that topic is sufficient.

Based on evidence from the previous trial of 1stClass@Number 1 (Nunes et al., 2018) it appears that these assumptions are accurate, but we planned to check these three assumptions again in the fidelity, quality, and compliance measures for this trial as it involves a different group of participants. In addition, there are also several assumptions relating to pupils' attitudes and behaviours that underlie the logic model outcomes. One of these is that increasing pupils' mathematical knowledge and conceptual understanding during the intervention (short-term outcomes) is sufficient to help them achieve success and feel more confident in maths (intermediate outcomes). Another assumption is that the combination of the small group setting, which gives pupils the opportunity to talk with each other about mathematics, and the 'special delivery' activities, which are designed to encourage mathematical talk with their family, will lead to improved peer interactions more generally. We thus explored, as part of the IPE, how pupils felt about the intervention, their motivation and confidence in maths, and the difference they felt the intervention has made for them more widely.

There is currently very limited evidence as to how much of a role parents/carers play in the success of the programme and to what extent the special delivery activities are used at home. In addition, one of the short-term outcomes in the logic model is that mathematical talk increases between parents/carers and their child. This assumes that the pupil has someone at home that they can talk to about school and specifically maths, but this may not be the case for all pupils. It also assumes that the special delivery activities will encourage parents/carers to use mathematical talk with their child(ren). There is potential for increased parent/carer engagement to benefit pupils' mathematical learning as outlined in the logic model, but there are also potential moderators on any effects. We therefore planned to investigate perceptions of the role of parents/carers in this intervention as an exploratory part of the IPE when considering the context, barriers, and facilitators.

Figure 1: Intervention logic model



Issues arising

Slow recruitment

Power calculations for this trial suggested that 235 schools needed to be recruited to ensure that the trial was adequately powered for FSM pupils. Recruitment started in February 2023 and was originally intended to be completed in July 2023. However, due to the requirements of the trial (for example, schools had to nominate at least ten pupils who were behind age-related expectations), staff shortages in schools, and the prevailing cost-of-living crisis, recruitment of schools was slower than anticipated. By the end of June 2023, 118 schools had been recruited with a further 41 in the recruitment pipeline. EHU, NFER, and the EEF agreed that 235 schools could not be recruited by July 2023 but that this number could be achieved if the recruitment period was extended to October 2023. It was, therefore, decided that the trial would be run in two cohorts. All training regions where sufficient numbers of schools had been recruited by July 2023 (a total of 156 schools) were in Cohort 1 and the remaining schools were in Cohort 2.

Sample size not achieved

By the end of the recruitment period in October 2023, EHU had recruited a total of 226 schools. The recruitment period could not be extended beyond October 2023 as this would not allow sufficient time for delivery within the same academic year to Cohort 2 schools. The requirement of 235 schools assumed a school attrition of 10% prior to randomisation; thus at design stage, 211 schools were expected to reach randomisation. All 156 schools in Cohort 1 had completed the baseline requirements and were randomised by October 2023. On this basis, we assumed that even with 10% attrition of the 70 Cohort 2 schools, a total of 219 schools (156 in Cohort 1 and 63 in Cohort 2) would be randomised, well above the expected 211 schools. Recruitment was stopped at this stage and the trial proceeded with 226 schools.

Impact of TA sickness on delivery and endpoint testing timelines for Cohort 1

Feedback from EHU's ECC trainers and from Cohort 1 schools to NFER test administrators suggested that TA sickness and other factors in schools had slowed the pace of delivery in some schools. As a result, pupils in some schools were not expected to receive the full intervention before endpoint testing in March 2024 (the endpoint testing window for Cohort 1 was 4–22 March 2024). In order to ensure that as much of the intervention as possible was delivered ahead of endpoint testing, Cohort 1 schools were asked to consider their progress with delivery and delay their endpoint testing dates to later in the testing window, if required, directly with the test administrators. The endpoint testing window for Cohort 1 schools was also extended to the week of March 25 provided both schools and test administrators were available during this week. For Cohort 2, the trial timetable was a bit more flexible and had two additional weeks built in for delivery to be completed. Additionally, there was no scope to extend the endpoint testing window for Cohort 2 without significantly disrupting the project timelines. Therefore, all Cohort 2 schools were asked to consider where they were with delivery of the intervention and schedule their endpoint testing dates on this basis.

Evaluation objectives

The research questions for the impact evaluation and IPE are provided below. Further detail on the evaluation design can be found in the study protocol (Lord *et al.*, 2023) and the Statistical Analysis Plan (Lord, Morton and Sahasranaman, n.d).

Impact research questions

Primary research question

RQ1 What is the difference in maths attainment measured by the Quantitative Reasoning Test (Nunes et al., 2015) of pupils receiving the 1st Class@Number 1 intervention in comparison to control pupils who do not receive the intervention?

Secondary research questions

RQ2 What is the difference in maths attainment measured by the Quantitative Reasoning Test of FSM-eligible pupils receiving the 1st Class@Number 1 intervention in comparison to FSM-eligible control pupils who do not receive the intervention?

RQ3 Is the 1st Class@Number 1 intervention as effective for pupils with Special Education Needs (SEN) as for pupils without SEN as assessed by the Quantitative Reasoning Test?

RQ4 Is the 1st Class@Number 1 intervention as effective for pupils with lower prior attainment as for pupils with higher prior attainment as assessed by the Quantitative Reasoning Test?

IPE research questions

IPE RQ1 Who was selected to participate in the intervention and how were they chosen?

IPE RQ2 To what extent was fidelity to (a) the intended training design and (b) the intended intervention delivery achieved?

IPE RQ3 How effectively was the intervention implemented in schools and what supported or hindered successful implementation?

IPE RQ4 What were the perceived benefits of the intervention for TAs and pupils?

IPE RQ5 What was business as usual (BAU) and did this change during the intervention?

IPE RQ6 What are the facilitators, barriers, and features of delivering at scale and is the intervention sustainable?

IPE RQ7 How manageable was it for schools to facilitate the pupil-randomised design?

Ethics and trial registration

An ethical review was undertaken as part of the National Foundation for Educational Research (NFER) start-up meeting in January 2023 where consideration was given to consent and the impact of the research on trial participants (pupils and practitioners). The evaluation was conducted in accordance with the NFER Code of Practice.

Each participating school's headteacher provided their agreement to participate in the trial by signing the Memorandum of Understanding (MoU—see Further Appendices)[8] that outlines the responsibilities of all parties involved in the trial. NFER shared a parent letter and withdrawal form with schools to be sent to parents/carers of all pupils that schools intended to nominate for screening (see Further Appendices). Through the withdrawal form, parents/carers had the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial.

A separate opt-in consent process was used for the pupil focus groups and applied only to those selected to participate. Given that pupils participating in this study are only six to seven years-old, we could not assume that all pupils would have the capacity to provide fully informed consent to participate. In addition, as the focus groups involved audio recordings it was especially important to ensure that parents/carers had the option to specifically consent to their child participating in this evaluation activity. We therefore provide parents/carers with a written information sheet about the focus groups containing full details about the focus group and what their child would be asked to do. Parents/carers were then asked to provide written opt-in consent of their willingness for their child to be invited to participate in the focus group by returning a consent form to the school, which passed this information on to the research team.

Pupil participation in the focus groups was voluntary, therefore even if a parent/carer had given consent for their child to participant, their child could still choose not to take part. Age-appropriate information about the focus groups was provided to pupils at the same time as parents/carers received information about the focus groups to allow them to discuss participation together. The researchers also read this information to pupils at the beginning of the focus groups

[8] All recruitment documents in the appendices are for Cohort 1. Cohort 2 documents were identical with the exception of the timetable.

to ensure pupils understood it and had the chance to ask any questions. If at this point a pupil decided that they would prefer not to participate, then they were able to return to their class. Prior to beginning the focus group, the researchers agreed some ground rules for the group with the pupils and had a discussion with them about the types of scenarios in which we would need to break confidentiality, to ensure they fully understood what this means.

The trial was designed, conducted, and reported to Consolidated Standards of Reporting Trials (CONSORT) standards and registered on the International Standard Randomised Controlled Trial Number (ISRCTN) registry (https://doi.org/10.1186/ISRCTN14875632).

Data protection

Personal data was processed as part of this trial. All data gathered was held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and treated in the strictest confidence by NFER, EHU and the EEF. NFER and EHU are joint data controllers for the duration of this trial up until the data is passed to the EEF archive.

The legal basis for processing personal data is covered by GDPR Article 6 (1) (f), which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils one of NFER's core business purposes (undertaking research, evaluation, and information activities) and it has broader societal benefits. Therefore, it is in our legitimate interest to process and analyse personal data for the administration of this RCT.

Prior to any sharing of data, NFER and EHU signed a sharing agreement governing the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, EHU provided an MoU to schools explaining the nature of the data being requested of schools, teachers, and pupils, how it would be collected, and how it would be passed to and shared with NFER. An information sheet and withdrawal form (provided by NFER) were also distributed by schools to parents of children they intended to nominate for the study before any of their child's data was transferred to NFER. Two separate privacy notices are available: one for schools and another one for parents (see Further Appendices).

For the purposes of the trial, EHU collected the names, roles, and contact details of a key contact person and the person signing the MoU when schools are recruited. It shared this data with NFER, which then contacted the key project contact person at participating schools to initiate data collection for the trial. NFER asked participating schools to nominate 12 to 16 eligible pupils for screening in addition to one TA and one Link Teacher who would support with intervention delivery. NFER shared the names, contact details, and job roles of the TAs and Link Teachers with EHU to enable it to coordinate the training sessions. EHU collected attendance data for TAs and Link Teachers at each training session and shared this data with NFER to create compliance measures that were linked to pupil-level data. EHU also shared the names and contact details of 1stClass@Number 1 trainers with NFER so that NFER could contact them to attend and observe a sample of training sessions. All personal data was shared via secure, password-protected data-sharing portals. NFER also collected pupil data from schools including names, date of birth, Unique Pupil Number (UPN), and FSM eligibility status for 12 to 16 eligible pupils who were subsequently screened using the SENT-R. The SENT-R and baseline QRT scores were also collected for these pupils. In addition to this data, for the eight pupils finally selected to participate, NFER also collected the endpoint QRT scores, details of any other interventions received, class name, and attendance at intervention sessions (if allocated to the intervention group).

For these pupils, background data including gender, FSM eligibility status, and SEN status were collected from the National Pupil Database (NPD). To obtain the information from the NPD, NFER provided the data sharing team at the DfE with the names of the pupils, their dates of birth, and UPNs, allowing a match to the NPD. After the matching process had taken place, the NFER then analysed this data using the Secure Research Service (SRS) managed by the Office for National Statistics (ONS). NFER accessed the data for analysis remotely through the SRS secure online system. The SRS does not allow users to remove or copy data from its servers.

As part of the IPE, NFER conducted online surveys of TAs and Link Teachers, observations of training sessions, interviews with trainers, observation of lesson delivery by TAs, and interviews with TAs and Link Teachers. NFER may share summaries of pseudonymised data from TAs' or Link Teachers' interviews and aggregated anonymised data from the TA and Link Teacher surveys with EHU to support their research. A small number of schools were also invited to participate in pupil focus groups. All NFER staff visiting schools had up-to-date DBS checks. All data gathered during interviews was stored securely.

At the end of EEF evaluations, all data is archived to allow for further secondary analysis. At this point, the EEF becomes the data controller and NFER is no longer responsible for the data and is no longer a data controller. Three months after the completion of the study all of the matched data (that is, to the NPD) will be added to the EEF archive and 'de-identified' before being made available to researchers. The EEF archive is hosted by the ONS and managed by the Fischer Family Trust (FFT). Other research teams may use the de-identified data as part of subsequent research through the Approved Researcher Scheme. NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication, all personal data will be securely deleted.

Project team

Name	Organisation	Role and Responsibilities
Pippa Lord	National Foundation for Educational Research (NFER)	Project director—responsible for overall delivery of the trial, IPE director.
Aarti Sahasranaman	NFER	Project manager—day-to-day management of the trial and delivery of the overall trial design.
Emma Moore	NFER	IPE lead—design and delivery of the IPE.
Eleanor Bradley	NFER	IPE researcher—IPE data collection.
Chris Morton	NFER	Statistician.
Kathryn Hurd	NFER	Research operations lead—test and schools administration lead, responsible for overseeing recruitment, school communications strategy, and testing.
Jishi Jose	NFER	Operations manager—day-to-day operations including preparation of recruitment documents, coordinating data collection, and point of contact for schools participating in the trial.
Lydia Wallis	NFER	Operations researcher—support project manager in delivery of data collection and school communications strategy
Yvonne Panteli	Edge Hill University (EHU)	Project manager—day-to-day operations including coordination of school recruitment, set up of programme delivery.
Lisa Leach	EHU	Every Child Counts development lead—school recruitment and intervention delivery plus school visits.
Linda Lavagna-Slater	EHU	Every Child Counts development lead—school recruitment and intervention delivery plus school visits.
Jonathan Glazzard	EHU	Contributed at project set-up, not involved with project from July 2023.
Jane Moore	EHU	Dean of Faculty of Education—responsible for the overall set up and delivery for the intervention.
Peter Hick	EHU	Principal Investigator for Edge Hill—responsibility for providing academic oversight on behalf of the University and leading on a parallel project on teaching assistant professional development.

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-arm, multi-site trial (i.e., within school randomisation at pupil level)
Unit of randomisation		Pupil
Stratification variable(s) (if applicable)		Pupil FSM-eligibility status; school
Primary outcome	Variable	Maths attainment
	Measure (instrument, scale, source)	(Nunes *et al.*, 2015)
Secondary outcome(s)	Variable(s)	NA (there were no secondary outcomes for this trial)
	Measure(s) (instrument, scale, source)	NA
Baseline for primary outcome	Variable	Maths attainment
	Measure (instrument, scale, source)	(Nunes *et al.*, 2015)
Baseline for secondary outcome(s)	Variable	NA (There were no secondary outcomes for this trial)
	Measure (instrument, scale, source)	NA

This is an effectiveness trial to assess the impact of the 1stClass@Number 1 intervention on the maths attainment of Year 2 pupils. We were particularly interested in the impact on disadvantaged (FSM) pupils and so the trial was powered to detect the impact on this subgroup.

The 1stClass@Number 1 intervention is offered to small groups of four pupils and randomisation was within schools at the pupil level. The decision to randomise at the pupil level was made following practical considerations of running a very large school-randomised trial. Assuming a conservative pre test/post test correlation of 0.29 as determined in the previous effectiveness trial (Nunes et al., 2018), our power calculations for a school-randomised trial powered for FSM-eligible pupils returned a sample size of 550 schools (assuming 50% of pupils selected to the trial are FSM-eligible). Through conversations with EHU, we established that these numbers far exceeded both their recruitment and delivery capacity if the trial were to be run for a single cohort during the academic year. We explored the possibility of running the trial in two cohorts over two academic years to address these recruitment and delivery challenges but this option proved to be financially prohibitive for such a large number of schools. Having explored these options, we ultimately decided to pursue a pupil-randomised design within schools as working with the same number of schools in this design will give us more power when compared to a school-randomised design. Powering the pupil-randomised trial for FSM-eligible pupils returned a sample size of 235 schools.

Based on the above, EHU planned to recruit 235 schools. Recruitment of schools was originally scheduled to take place between February and June 2023. As recruitment progressed, it became evident that it would be difficult to reach the recruitment target of 235 schools by end of June 2023. Schools noted numerous challenges including the ongoing cost of living crisis, uncertainty over staffing, and lack of budget that prevented them from signing up to participate in the trial. To ensure that the original target of 235 schools was recruited so that the trial could continue to be powered for FSM pupils, the EEF, NFER and EHU agreed that a split cohort design would be the most feasible solution. In this revised design, the trial was run in the same academic year (2023/2024) with the 235 schools split into two cohorts. Since recruitment by EHU took place by training regions, training regions were allocated to either Cohort 1 or Cohort 2 depending on their likelihood of being filled earlier or later, respectively. Furthermore, due to challenges with recruitment of schools to Cohort 2, a small group of schools were recruited to an online training group to support recruitment targets for Cohort 2. Schools in Cohort 1 started receiving training and delivering the intervention in November 2023; for Cohort 2 schools it was January 2024.

Eight pupils from within each school were randomised on a 1:1 basis into two arms: intervention and control. Given the focus on FSM pupils in this trial, randomisation was stratified by pupils' FSM eligibility to ensure a balance of FSM pupils in both arms. Pupils assigned to the intervention arm received the 30-session 1stClass@Number 1 intervention in addition to normal maths lessons. Pupils assigned to the control arm did not receive the intervention but were able to access other support provided to them as a matter of course. The evaluation collected information about the maths interventions (if any) that trial pupils received.

The evaluation measured the impact of 1st Class@ Number 1 on the maths attainment of Year 2 pupils. Consistent with the previous effectiveness trial, the primary outcome measure used was the Quantitative Reasoning Test (QRT). There were no other attainment outcome measures (see Secondary Outcomes section for detail).

All schools paid a subsidised training fee of £200 (the usual cost is £1,100) and were eligible to receive an incentive payment of £500 upon completing all trial-related activities.

Participant selection

School eligibility

The eligibility criteria for schools to participate in this trial were:

- infant or junior school with children in Year 2, as of 1 September 2023;
- *not* taking part in the EEF-funded programmes ARK Mathematics Mastery or Mathematical Reasoning in 2023/2024;
- *not* having a TA currently working in the school who has already completed any 1stClass@Number 1 training;
- *not* delivered 1stClass@Number 1 after 1 September 2019;
- should nominate a TA and a Link Teacher who are able to access face to face training in one of the 16 training venues; and
- pay the subsidised training cost of £200 (usual cost is £1,100).

Note, as a requirement for projects funded by the Accelerator Fund, EHU also had a target for half of the recruited schools to be in Education Investment Areas (EIAs). In addition, because the trial required at least 50% of the pupils to be FSM eligible, EHU targeted schools with high proportions (more than 30%) of such pupils during recruitment. To ensure that the trial included a sufficient proportion of FSM pupils, EHU monitored the percentage of FSM pupils in Year 1 in schools at the expression of interest (EoI) and MoU stages, and NFER checked the school-level data received at MoU stage and subsequently the pupil-level data at each stage described below.

In addition, schools were requested to ideally have 12 to 16 pupils suitable for nomination. Note that at the EoI stage schools were asked to indicate whether they had at least ten pupils whom they considered to be suitable for the trial (12 to 16 was the ideal, but at least ten was in the EoI).

Pupil eligibility

To identify pupils eligible to participate, we followed a two-step process. First, schools were asked to nominate 12 to 16 pupils whom they believed were eligible to receive 1st Class@Number 1; EHU provided guidance to schools on how to identify such pupils (see below). This pupil nomination step took place in the summer term of 2023 when pupils were in Year 1 and for some Cohort 2 schools took place in the autumn term of 2023 when pupils were in Year 2. In the second step (September to November 2023), these nominated pupils were screened using the SENT-R so that NFER could identify the final eight pupils who would be selected to participate in this trial (see Selection Process below for further details). Pupil selection took place in the autumn term of 2023 when pupils were in Year 2. It was our expectation that not all pupils who were nominated for screening would be eligible based on their SENT-R scores. This is why we asked schools to nominate more pupils than the final eight who would be selected to the trial.

Pupil nomination for screening

There were three broad eligibility criteria to determine which pupils could be nominated for screening:

1. The pupil should have been achieving at around 12 months behind age-related expectations in maths towards the end of Year 1.
2. The pupil should not have already been identified to participate in another mathematics intervention when 1st Class@Number 1 is being delivered (autumn 2023/spring 2024).
3. This trial has a focus on the impact of 1st Class@Number 1 on disadvantaged pupils, especially those in receipt of free school meals. Schools were, therefore, encouraged to nominate FSM-eligible pupils who met the above criteria.

Pupils who had fallen behind their peers and require additional support would typically:

- be able to count forwards in ones to ten;
- have some knowledge of number facts and also some understanding of the composition of number within ten;
- be able to perform simple addition and subtraction calculations using ‘counting all’ and ‘counting out’ approaches;
- be able to read and write numbers but lack secure understanding of their magnitude and quantity;
- struggle with mathematical vocabulary; and
- lack confidence in mathematics and be reluctant to talk about their mathematical learning.

NFER and EHU developed a guidance document for schools to enable them to choose the 12 to 16 most appropriate pupils for screening (see Further Appendices). The document also outlined how schools should proceed with pupil nomination if they had fewer than 12 pupils to nominate. When schools were only able to nominate eight pupils who met the criteria above, our recommendation was that they nominate two additional pupils to accommodate any change in circumstances of the eight pupils thought to be eligible for screening. While these remaining two pupils may not fully meet the selection criteria above, they must be expected to benefit from the intervention. This guidance document was shared by EHU with schools after they signed the MoU.

Pupil selection for trial

Once schools nominated pupils for screening and shared their administrative data with NFER, the next step was to select the final eight pupils for the trial. Nominated pupils from each school undertook the SENT-R (and QRT, see Baseline Measures section below) from September to November 2023 at the start of Year 2 to objectively confirm that they were achieving below age-related expectations. Schools were asked to administer Form A of the SENT-R (the test

kit provides two versions, Forms A and B to allow for pre- and post- measures).[9] The SENT-R screening test was administered 1:1 in schools by teachers and the test papers were sent back to NFER for processing the scores. NFER shared a customised report with each school with each pupil's SENT-R scores and recommended eight pupils to take part in the trial. Pupils were prioritised for selection according to their SENT-R score and FSM status. Pupils were selected by NFER according to priority groups, which were (ordered by highest to lowest priority):

i. FSM pupils with SENT-R scores of 49 or below;
ii. non-FSM pupils with SENT-R scores of 40 or below;
iii. non-FSM pupils with SENT-R scores of 41–49; then
iv. any pupils with SENT-R scores above 49.

The eight pupils for randomisation were first selected from group (i). If additional pupils were needed to reach eight, they were taken from group (ii), and so forth. While pupils scoring 40 or below are considered to be around 12 months behind age-related expectations, in agreement with EHU we also allowed the selection of pupils with a raw score between 41 and 49, where a raw score of 49 equates to a number age of 6:10, i.e., 6 years and 10 months (this is part of usual practice that EHU experiences). Schools were asked to briefly confirm the recommended shortlist and also given the option of replacing recommended non-FSM pupils with others who had not been shortlisted by NFER but whose screening results still met all other criteria set out above. While schools were encouraged to not replace the recommended FSM pupils, there were some instances where this replacement took place (see IPE section for details).

This selection mechanism allowed all recruited schools to continue participating in the trial even if they did not have eight pupils who strictly met the SENT-R eligibility criteria. This approach also ensured that schools had visibility of, and some input into, the final confirmed list, allowing them to consider factors such as group dynamics. It is important to note that the randomisation was undertaken independently, and as would be expected in an RCT: schools had no say in the actual group allocation. The Subgroup Analyses section below provides details on how we managed the SENT-R range in our dataset since some pupils fell above the raw score cut-off. An alternative approach we considered was that of unequal randomisation if fewer than eight pupils were eligible based on the strict SENT-R cut-off score of 40. However, after considering the analytical challenges arising from consistently allocating more pupils to the intervention arm (as the intervention requires delivery to groups of four pupils), we ultimately decided against this approach. This option was permitted only in a small number of cases where fewer than eight pupils in a school completed the SENT-R screening test (see the statistical analysis plan and Randomisation section for further details).

Recruitment and pupil data collection

EHU, the intervention delivery partner, along with their local ECC trainers were responsible for recruiting schools for this trial. They did so through direct contact with schools by EHU ECC trainers, media engagement, social media channels such as X, and other promotional events. The EEF and NFER supported their recruitment efforts by promoting the trial through their newsletters and social media channels. Interested schools completed an online EoI to help EHU ascertain schools' eligibility to participate in the trial. Eligible schools were then sent the school information sheet and MoU (see Further Appendices for all trial recruitment documents). Schools signed up to the trial by the headteacher signing the MoU and providing the name of a key project contact who acted as the coordinator of the trial in the school. Once a school signed up, EHU shared the school's name and details of the headteacher and key project contact with NFER. NFER then contacted the key project contact to initiate data collection and other trial-related activities. At this stage, where schools did not have 12 to 16 eligible pupils, they were asked to nominate at least ten eligible pupils for the trial. If schools only had eight eligible pupils that they could nominate, they were asked to nominate an additional two (that is a total of ten pupils) who would benefit from the intervention. This was done to help us accommodate any change in circumstances of pupils so that eight pupils were ultimately selected to participate. Overall, EHU recruited 226 schools of which 156 were allocated to Cohort 1 and 70 to Cohort 2. While the sample size calculations at design stage suggested a recruitment target of 235 schools (including 10% school attrition between recruitment and randomisation),

[9] Form A and Form B are parallel forms of the SENT-R test assessing the same core numeracy skills but with different question sets. This allows for multiple assessments using different questions to avoid test fatigue when the test is administered more than once. In the context of this trial, the SENT-R was only used as the screening test at baseline. To ensure consistency, all schools were asked to administer Form A to pupils for screening.

it was decided that the trial could proceed with slightly fewer schools as long as we could monitor school attrition closely and ensure that at least 211 schools reached the randomisation stage (see Sample Size section below). None of the schools recruited to the trial dropped out before randomisation ensuring that we met this requirement.

contacted schools after they signed the MoU to share the parent letter and withdrawal form and the pupil data form. Schools were asked to share the parent letter and withdrawal form with the parents/carers of all pupils that they intended to nominate for screening. At this stage, parents/carers had the opportunity to withdraw their child from the evaluation and associated data processing. Once the withdrawals were processed, schools shared pupil administrative data for the remaining pupils with NFER via the pupil data form. This included pupil names, date of birth, unique pupil number (UPN), FSM eligibility, and class name in Year 2 (if known). Schools were also asked to share details of the TA and Link Teacher (names, role, email address, telephone number) they wished to nominate for training. Schools that had not decided which staff members to nominate for training at this stage were given another opportunity in autumn 2023 to provide these details.

In total, 226 schools nominated 2,833 pupils for screening of whom 2,625 completed the SENT-R screening test. Once the screening and baseline assessment were complete, schools confirmed the list of eight selected pupils with NFER. Completion of the screening and confirming details of the eight selected pupils were prerequisites for randomisation. Independent random assignment of the eight pupils to the intervention or control arms in a 1:1 ratio took place in October 2023 (Cohort 1) and January 2024 (Cohort 2). Details of which arm each pupil was randomly assigned to were shared with schools to enable them to deliver the intervention to the appropriate pupils. A total of 1,797 pupils were randomised, 904 to the intervention arm and 893 to the control arm.

To minimise contamination between intervention and control pupils, TAs were provided with 'dos and don'ts' guidance (see Further Appendices) providing them with practical suggestions on avoiding exposure of control pupils to the pedagogical approaches of 1stClass@Number 1. For example, TAs were asked to avoid exposure of control pupils to displays of intervention pupils' work.

Outcome measures

Baseline measures

We used the Quantitative Reasoning Test (QRT) score (Nunes et al., 2015) in its entirety as the baseline measure. At baseline, the QRT was administered by teachers in September 2023 (Cohort 1) and October to November 2023 (Cohort 2) and the test papers were sent back to NFER for marking. This baseline testing window coincided with the SENT-R screening window. To minimise pupil fatigue, teachers were instructed to first administer the QRT to all nominated pupils in a group. Administration of the QRT to a group of pupils was expected to take 30 to 45 minutes to allow all pupils sufficient time to attempt each question. After a gap of two to three days, they were asked to administer the 1:1 SENT-R individually to each pupil. Teachers were provided detailed guidance by NFER for the administration of the QRT and SENT-R. Although the baseline QRT tests were completed before randomisation, the tests were marked after randomisation was completed. The tests were marked by trained NFER staff who were blinded to the pupils' treatment allocation. Baseline QRT scores were included in the models as a covariate. This was described in detail in the study's statistical analysis plan (SAP) (Lord, Morton and Sahasranaman, n.d).

Primary outcome

In line with the previous effectiveness trial, the primary outcome for this trial was maths attainment as measured by the QRT (Nunes et al., 2015). The QRT comprises four scales that measure additive composition, the inverse relationship between addition and subtraction, additive reasoning, and multiplicative reasoning. The scales comprised in the test were deemed as a good match for the 1stClass@Number 1 intervention, which aims to build pupils' confidence in number sense, place value, and calculation. The reliability and construct validity of the QRT are demonstrated by the good alpha reliability levels of the four scales and test-retest correlation for the total score ($R = 0.78$, $p < 0.001$: Nunes et al., 2015). The four scales correlate significantly with each other, and each measure has a factor loading above 0.8. The QRT has been validated among young children aged five to six years and has been shown to be a strong predictor of KS1 attainment (correlation of 0.7 when the assessments were 14 months apart: Nunes et al., 2007). Based on the logic model for 1stClass@Number 1, the intervention is expected to increase maths attainment. Given the relevance of the

QRT to the topics covered in the intervention and its psychometric properties, we believe this outcome measure is suitable to address the research questions about the impact of 1stClass@Number 1 on the maths attainment of pupils.

At endpoint, March 2024 for Cohort 1 and May to June 2024 for Cohort 2, the QRT was administered in schools by NFER test administrators who were blinded to treatment allocation. The tests were marked by trained NFER staff who were also blinded to the pupils' randomised assignment. All resources required to train NFER test administrators (and teachers at baseline) and to administer and score the tests were provided to the research team by the developers of the QRT.

Secondary outcomes

There were no secondary outcome measures for this trial. The previous trial used Key Stage 1 attainment as a secondary outcome, however, this was not possible in this current trial. Starting summer 2024 this test was no longer mandatory in schools. In addition, the trial schools had the option to deliver 1stClass@Number 1 to control pupils following the trial period but before any longer-term attainment outcome measurement would be possible.

Sample size

All sample size calculations were performed in the PowerUp software (Dong and Maynard, 2013) using the sheet `BIRA2_1c` for block randomised trials with a constant intervention effect across blocks (in this case schools).

We note that `the sheet 'BIRA2_1r'`, for block randomised trials with a random intervention effect, more closely reflects the models used in this evaluation than the calculation in `BIRA2_1c`. The calculation in 'BIRA2_1r' requires a parameter ω, which is the proportion of school-level variance that is due to the impact of the intervention varying between schools. At the design stage we were unsure of a realistic value for ω, which varies considerably between studies, as described by Hedges and Rhoads (2010). The primary analysis model effectively assumes $\omega = 0$ due to the absence of random slopes which, combined with the unconditional ICC of 0.35 observed for this evaluation, suggests an MDES of 0.10. However, the sensitivity analysis including random slopes estimates that $\omega = 0.22$, resulting in a slightly higher MDES of 0.11. In either case, the primary analysis MDES of 0.13 seen in Table 8 is likely to be somewhat conservative and future studies with a similar design should use the `BIRA2_1r` calculation, possibly basing their estimate of ω on that reported here.

Protocol stage

A central question for this evaluation is what impact 1stClass@Number 1 has among FSM pupils. However, in line with the EEF's remit across all trials, the primary research question in this trial assessed the impact on all pupils. This trial was powered to detect an impact for all pupils and also for FSM pupils, respectively, with a false-positive rate of 0.05 for each of these hypotheses. No multiple testing correction was applied to restrict the family-wise error rate to 0.05 if these hypotheses are considered simultaneously. The sample size was specified so that the evaluation would be powered for a minimum detectable effect size (MDES) of 0.2 in the FSM-eligible subgroup. The MDES of 0.14 among all pupils followed from this sample size specification.

The recruitment target for this study was 235 schools with the anticipation that approximately 211 would be randomised. This is because, based on previous NFER evaluations, we predicted approximately 10% of recruited schools would not reach the analysis stage and that this attrition would mostly occur before randomisation. As eight pupils would be randomised per school, 1,688 pupils randomised were predicted, split evenly between the intervention and control. It was assumed for the power calculations that 50% of pupils in total across all schools would be eligible for FSM (see Participant Selection section above for the strategies employed to support this), which would mean 844 FSM pupils randomised. Previous NFER evaluations suggest approximately 15% of randomised pupils would not be included in the primary analysis, which meant 1,432 pupils (716 eligible for FSM) would be analysed. Assuming a pre-test/post-test correlation of 0.3, this sample size was sufficient to detect an effect size of 0.2 among FSM-eligible pupils (or 0.14 amongst all pupils), with a power of 0.8.

The pre test/post test correlation of 0.3 predicted for this study, which was lower than for many EEF trials, was based on the previous evaluation of 1stClass@Number 1 (Nunes et al., 2018), which observed a pre test/post test correlation of 0.29. For the current evaluation, we took the additional step of screening pupils using a different instrument (the

SENT-R) to that used for the baseline measurement (the QRT), which we believed could improve the pre/post correlation. Nevertheless, we made the conservative assumption at the protocol stage that the pre/post correlation would be similar to that observed in the previous evaluation.

Randomisation stage

Two hundred and twenty-six schools (1,797 pupils) were included in randomisation, which was more than expected at the protocol stage. The randomisation-stage MDES was calculated based on all 226 schools, with no school attrition between randomisation and analysis assumed. The 1,797 randomised pupils were reduced by 15% in the MDES calculation to represent the anticipated amount of pupil-level attrition. The proportion of pupils randomised that were FSM-eligible was 64% (this was the same in both cohorts), which was also higher than anticipated. As a result, the updated MDES at the randomisation stage was slightly lower for all pupils (0.13) and substantially lower for FSM pupils (0.16).

Analysis stage

Only two schools withdrew from the trial between the randomisation and analysis stages, so data from 224 schools (1,560 pupils) was included in the primary analysis. The proportion of FSM pupils, which was now defined using the NPD variable `EVERFSM_6_P` rather than being provided by schools, remained at 64% (993 pupils). As anticipated, the pre test/post test correlation was low compared to many EEF-funded evaluations but higher than the previous trial at 0.44 (0.39 for FSM pupils). The resulting MDES was 0.16 for FSM pupils and 0.13 for all pupils, the same as at randomisation and lower than in the protocol (see Table 8 in the Results section).

Randomisation

The eight pupils selected within each school were randomised in a 1:1 ratio as a block within the school, so that four were assigned to the intervention and four to the control. This was necessary to ensure both an equal randomisation ratio and that the intervention was delivered as intended, in tutoring groups of four. Randomisation occurred in three tranches: two for the first cohort of pupils[10] and one for the second cohort. This did not introduce additional analytical considerations (for example, the need to include cohort as a stratifier in analysis models) as randomisation was already blocked within each school and all pupils within a school were randomised in the same tranche.

The randomisation was also stratified by FSM eligibility within each school using FSM data collected from schools.[11] This stratification maximised statistical power for the FSM subgroup: for a given sample size, a 1:1 randomisation is superior to any other allocation ratio in terms of power.

The pupil selection process was designed so that exactly eight pupils were randomised at each school. However, in practice there were a small number of schools that were not able to provide eight pupils for randomisation (eight provided seven pupils, one provided five pupils). In these cases, four pupils were randomised to the intervention and the remainder to the control, ensuring that the intervention was delivered as intended to groups of four. This approach has a small chance of introducing a small amount of bias as it introduces an association between randomisation allocation and any factors causing less than eight pupils to be randomised (for example school size). We believe our approach to be the best one available but acknowledge that others could have been taken. For example, allowing all four intervention pupils to participate in sessions but discarding their data at random until there was an equal number of control and intervention pupils analysed at the school. This approach would have insured against bias but reduced sample size.

Randomisation was carried out by an NFER statistician using R code, which was stored for reproducibility and transparency. The statistician was not blinded to group allocation.

[10] Within Cohort 1, schools that were not able to return their SENT-R and baseline QRT data in time to be randomised in the first tranche were randomised in the second.

[11] For all 'pre-analysis' purposes (including recruitment, pupil selection, and stratified randomisation) the FSM variable was created by asking schools which of their pupils were currently eligible for FSM.

Statistical analysis

The impact analysis generally followed the EEF analysis guidance (EEF, 2022) although there is a notable divergence for the primary analysis model, which is discussed below. All analyses were conducted using version 4.4.0 of the R software (R Foundation, 2023b).

Mixed effects models with a random intercept only were analysed using R package `lme4` (Bates et al., 2015). Outputs from the model with a random intervention effect were obtained using the R package `eefAnalytics` (Uwimpuhwe et al., 2023) with outputs cross-checked against the same model calculated by `lme4`. `eefAnalytics` provides additional functionality that makes certain processes (for example, calculating effect sizes) more convenient, but the two packages agreed in their output when calculating the same statistics.

An intention-to-treat approach was followed throughout (except the compliance analysis), with pupils analysed according to their intervention or control group assignment, regardless of their degree of participation in the intervention. Analysis was conducted on complete cases only; pupils with any missing analysis variables were excluded. The missing data analysis investigated the sensitivity of results to this choice.

Primary analysis

There is a single primary research question for this trial:

RQ1 What is the difference in maths attainment measured by the Quantitative Reasoning Test of pupils receiving the 1st Class@Number 1 intervention in comparison to control pupils who do not receive the intervention?

As this is an effectiveness trial that uses a multisite design, the EEF Statistical Analysis Guidance indicates that the appropriate analysis to answer this research question is a mixed effects model with a random intervention effect per school, in addition to a random intercept. This would allow for the possibility that the impact of the intervention varies between schools. However, we thought it possible that such a model would face technical issues such as lack of convergence, given that there are only eight pupils randomised per school. This problem seemed likely to be exacerbated in the subgroup analyses below, which would include as few as one pupil per school. A random intervention effect was therefore not included in the primary analysis, although it was added as a sensitivity check (see Additional Analysis section). [12]

In the absence of a random intervention effect, the choice remained as to whether to represent schools using a random intercept (leading to a mixed effects model) or as a fixed covariate (leading to a fixed effects model). As this trial has close to a balanced design with a 1:1 allocation ratio within schools, [13] the two approaches should produce very similar effect estimates and standard errors, although the standard errors for the mixed effects model will be slightly smaller (Feaster, Mikulich-Gilbertson and Brincks, 2011). It was decided that a two-level (pupil and school) linear mixed effects model would be used for the primary analysis. A school-level random intercept was included, allowing for the within-school clustering of pupil error terms.

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \mathrm{intervention}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + b_j + \epsilon_{ij}$$

where:

QRT_{ij} = endpoint QRT score of pupil i in school j;

β_0 = intercept term;

[12] Our concerns were largely misplaced, as the sensitivity check models did in fact converge in practice. Despite this, the primary analysis still only has a random intercept, in accordance with the Statistical Analysis Plan.

[13] Due to missing data there will not quite be the same number of pupils analysed in each school. The randomisation allocation was also not quite 1:1 at some schools, as described in the randomisation section.

intervention_{ij} = indicator for whether pupil i in school j was randomised to the intervention (1) or control (0);

QRT_BL_{ij} = baseline QRT score of pupil i in school j;

FSM_{ij} = indicator for whether pupil i in school j is eligible for free school meals[14], yes (1) or no (0);

b_j = school-level error term (random intercept); and

ϵ_{ij} = pupil-level residual error term.

Here the estimate of β_1 represents the average impact of 1st Class@Number 1 on QRT scores with a 95% confidence interval for this estimate calculated using the profile likelihood. The estimate and confidence interval were converted to a standardised effect size, as described in Estimation of Effect Sizes below.

Analysis in the presence of non-compliance

'Compliance' was defined in terms of the number of intervention sessions attended by pupils and also on the number of training sessions attended by TAs (for compliance definition (3) below). The section Attendance Register and Intervention Records in the IPE methods outlines how this data was collected. Three binary compliance definitions were originally proposed in the SAP:

(1) Pupil attended all 30 sessions.

(2) Pupil attended at least five of the six sessions in each of the five topic areas, implying at least 25 of the 30 sessions attended in total.

(3) Pupil attended at least five of the six sessions in each of the five topic areas, as for (2) above. Additionally, the pupil's TA attended all four of the initial training sessions in some form (catch-up sessions, online or otherwise, were included).[15]

An instrumental variable analysis was performed using two-stage least squares methods (Angrist and Imbens, 1995) to estimate the effect of compliance with the intervention on endpoint QRT scores. Instrumental variable analysis provides a method for estimating the causal effect of compliance with the intervention, one that is robust to the presence of unobserved confounding variables. However, its validity depends on key assumptions, in particular that randomised intervention allocation can impact the QRT outcome solely through receipt of the intervention ('compliance'). This is known as the 'exclusion restriction'.

A key implication of the exclusion restriction is that when compliance is not truly dichotomous but is instead defined by applying a threshold to an underlying count (for example, the number of sessions attended), the intervention must have no effect below the chosen compliance threshold. For example, for compliance definition (1), attending fewer than 30 sessions would be assumed to provide no benefit to attainment, even if most sessions were attended. This assumption seems unlikely to hold in the context of the intervention, suggesting that the Complier Average Causal Effect (CACE) impact estimates obtained for compliance definitions (1) to (3) are upwardly biased.

Due to concerns around the violation of the exclusion restriction, in agreement with the EEF, a fourth compliance definition was added after publication of the SAP. The impact of compliance, as defined by (4), was not a primary research focus but served to provide additional context for definitions (1) to (3) in the case of exclusion restricted violation.

(4) Pupil attended at least four of the six sessions in at least four of the five topic areas, implying at least 16/30 sessions attended in total.

---

[14] `EVERFSM_6_P` from the 2023/2024 spring census was used to measure FSM eligibility in all analysis models. The variable indicates whether a pupil has been eligible for FSM at any point in the previous six years. For this evaluation of Year 2 pupils, FSM records do not go back six years since FSM status is first recorded in the reception year.

[15] Aspects of fidelity such as whether sessions were attended in person were still recorded and presented descriptively as part of the IPE analysis but were not included in these compliance measures.

Compliance definition (4) was agreed with EHU as a 'minimum' compliance measure—the threshold below which the intervention was judged to provide little or no benefit (that is, the highest threshold at which the exclusion restriction holds). As a result, the estimate for compliance definition (4) can therefore act as a lower bound for compliance definitions (1) to (3). For example, if the effect size for compliance definition (4) was 0.17 and for compliance definition (1) was 0.30, we could conclude that the unbiased effect size for compliance definition (1) falls between 0.17 and 0.30. This lower bound approach, which was proposed by Gerber and Green (2012), is not an entirely satisfactory solution as considerable uncertainty remains around the unbiased effect size. However, we consider it important to acknowledge this uncertainty. It is also worth noting that the minimum compliance measure provided by EHU was a suggestion based on its experience in managing the intervention rather than on rigorous empirical analysis.

The impact of compliance based on each definition was modelled for all pupils and for the FSM-eligible subgroup separately, resulting in a total of eight models. In the first stage of each model the compliance indicator was regressed on the intervention indicator, together with covariates from the primary analysis model (baseline QRT score, FSM status). This first stage linear regression was:

$$\mathrm{compliance}_{ij} = \beta_0 + \beta_1 \mathrm{intervention}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + \epsilon_{ij}$$

In each model the compliance indicator, $\mathrm{compliance}_{ij}$, was constructed based on one of definitions (1) to (4) above and generally took the value zero for control pupils.[16] For the second stage, endpoint QRT scores were regressed on each pupil's predicted compliance value, $\widehat{\mathrm{compliance}}_{ij}$, obtained from the first stage, in the following linear regression:

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \widehat{\mathrm{compliance}}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + \epsilon_{ij}$$

The coefficient for predicted compliance β_1 in this second stage is the CACE estimate for the effect of compliance on endpoint QRT scores.

Results from both regression stages are reported in the results section. All instrumental variable analyses were performed using the R package `ivreg` (Fox et al., 2021). These models do not include school-level random effects so the standard errors produced by `ivreg` by default are incorrect. Instead, cluster-robust standard errors were calculated using the R package `sandwich` (Zeileis, 2006; Zeileis, Köll and Graham, 2020).

Missing data analysis

The missing data analysis was performed for the primary analysis model. Here, the number of missing cases was 237, corresponding to 13.2% of randomised pupils. As this was higher than 5% (the threshold suggested in the EEF Statistical Analysis Guidance), there was potential for bias caused by missing data and further analysis was required.

As both baseline and endpoint QRT scores had missing data, a mixed effects logistic regression model with two levels (pupil and school) was conducted for each. The outcomes were the logit probabilities of the respective QRT scores being missing. All other variables from the primary analysis model were included as covariates, together with auxiliary variables potentially associated with missingness:

- pupil SEN in 2023/2024;
- pupil SENT-R score;
- pupil's school type (academy, maintained, or independent) in 2023/2024; and
- proportion of pupils eligible for FSM in 2023/2024 at the pupil's school.

These regressions were performed both for all pupils and the FSM subgroup. As explained in the SAP, the plan was to include any additional variables that demonstrated an association with missingness in the endpoint QRT score ($p < 0.05$) as covariates in the primary analysis model. If missingness of the baseline QRT score was associated with any of

[16] As this trial used within-school randomisation there was a possibility of control pupils taking part in the intervention, so data was collected on control pupil participation in 1stClass@Number 1 sessions. A small number of control pupils did in fact participate in some sessions, as described in Results section.

the additional variables listed above, a sensitivity analysis was planned using multiple imputation to address missing data. However, in practice, no variables were found to be associated with missingness in either QRT score (see Results section), so these further steps were not required.

As the available variables did not account for missingness in QRT scores, we performed further sensitivity analyses exploring the possibility that some endpoint QRT scores were 'missing not at random' (MNAR).[17] This analysis was not specified in the SAP. The missing data patterns that were considered are described in Table 4.

Table 4: Missing data patterns that were considered possible in the missing data sensitivity analysis

Missing data pattern	How missing outcome data was imputed for intervention pupils	How missing outcome data was imputed for control pupils	Potential to bias primary analysis result
MAR	Imputed in the usual manner.*	Imputed in the usual manner.*	None: a complete case analysis will be unbiased.
MNAR balanced	Each imputation reduced by one standard deviation.**	Each imputation reduced by one standard deviation.**	Medium: can cause bias if there is a different attrition rate between intervention and control pupils.
MNAR unbalanced	Imputed as-is if they're a control pupil (that is no benefit received from intervention)***	Imputed in the usual manner.*	High: will bias the primary analysis result.

*That is, outcomes were imputed using the multiple imputation process described below, without additional adjustments to represent a MNAR scenario being made.
**The choice of one standard deviation lower (based on all observed endpoint QRT scores) is somewhat arbitrary; it is intended to be substantially lower but still realistic as it is within the range of the observed data.
***When the intervention pupil's outcome was imputed as part of the multiple imputation procedure the intervention indicator was used to predict the imputed outcome; this indicator was set to 'control' (0). Of course, when models are later analysed and pooled these pupils will be analysed as intervention pupils.

Instead of assuming that one missing data pattern from Table 4 applied to all pupils, it was possible to refine our inference based on the missing endpoint test code collected for each pupil. There were six possible codes, which are given in the first column of Table 5. For each code we considered what missing data pattern represented a realistic 'worst case' scenario (in terms of potential to bias the primary analysis) for pupils with that code. Outcomes for each pupil were then imputed accordingly (as given in Table 4). This is 'scenario 1' in the Results section; two more optimistic scenarios (scenarios 2 and 3) were also considered.

Table 5: The six possible reasons why pupils had a missing QRT outcome and what we considered to be the reasonable worst case missing data pattern for those pupils

Reason for missing QRT outcome	Missing data pattern in reasonable worst-case scenario	Explanation
School withdrawal	MNAR unbalanced	School withdrawal may indicate lack of resources (for example staff time) to implement the intervention properly or lack of engagement with the intervention.
Pupil left school	MNAR unbalanced	Intervention pupils that leave the school during the trial period will miss some or all intervention sessions (effectively non-compliance).
Pupil did not want to take test	MNAR balanced	Reluctance to take test could indicate lack of confidence and/or knowledge of test material.

[17] MNAR scenarios were not explored for missing baseline QRT scores. Baseline QRT scores were less often missing than endpoint QRT scores and were measured before randomisation, so there is no clear mechanism by which missing data would bias the intervention estimate.

Pupil present but excluded from the test*	MNAR balanced	Poor behaviour may be correlated with lower attainment. It may also be an expression of reluctance to take the test, as in the box above.
Pupil absent on day of test	MAR	No reason to believe that absence on the day of the test is correlated with worse outcomes (except possibly via persistent absence, which is addressed in the missing data results).
Pupil withdrawal*	MAR	Withdrawals were generally made by parents relatively early in the trial period, presumably due to concerns around privacy or their child's welfare. There does not seem to be any reason to believe this would correlate with attainment.

*For these missingness reasons it is less clear what missing data assumptions are reasonable, but we note that very few pupils had these missingness reasons (see missing data results) so it should not substantially influence this sensitivity analysis.

Missing data was imputed using a two-level normal model with the `2l.lmer` function in the R package `mice` (Buuren and Groothuis-Oudshoorn, 2011). The imputation model included all variables from the primary analysis. Missing QRT outcomes were imputed stochastically in the usual manner, but with the imputations altered as described in Table 4 to represent them being MNAR. Missing QRT baseline and FSM data was imputed in the usual manner, without alterations to represent MNAR scenarios. Ten datasets were generated, each using ten iterations, and the primary analysis model was then re-run on each. Estimates from each model were pooled into a single set of estimates and standard errors that were compared to the primary analysis result. This entire procedure described in this paragraph was performed three times, under different sets of MNAR assumptions (scenarios 1 to 3, see the missing data results).

Subgroup analyses

Measuring the impact of 1stClass@Number 1 for FSM pupils is of particular importance given the lesser impact observed for this group compared to all pupils in the previous effectiveness trial (Nunes et al., 2018). This evaluation was therefore powered to detect an effect in the FSM subgroup. The primary analysis model was re-run for (i) the subgroup of FSM-eligible pupils and (ii) the subgroup of pupils not eligible for FSM,[18] using the `EVERFSM_6_P` variable from the 2023/2024 spring school census. The model was:

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \mathrm{intervention}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + b_j + \epsilon_{ij}$$

Subgroup analysis was also performed to determine the effect of the intervention for pupils with special educational needs (SEN), defined by converting the `SENprovisionMajor` variable gathered from the NPD into a binary indicator for SEN (1) or no SEN (0). This followed the same specification as the primary analysis, but only included pupils with SEN.

When selecting pupils for the trial, those with SENT-R scores of 40 or below were prioritised. However, to ensure that eight pupils were randomised per school—allowing the intervention to be delivered as intended in groups of four—224 pupils with higher scores were also included (209 pupils had scores of 41 to 49 and 15 pupils had scores of 50 plus). As 1stClass@Number 1 is intended for pupils with moderate difficulties in maths, it was important to assess whether the inclusion of these higher-attaining pupils affected the primary analysis estimate. To investigate this, a subgroup analysis was conducted focusing on pupils with lower SENT-R scores using a score threshold of 40 and below to define the subgroup.

For the primary analysis, data for pupils in Cohorts 1 and 2 were analysed together. As the two cohorts receive the intervention roughly a term apart, there was a possibility that the effect of the intervention differed between them (for example, due to them being at different stages in their maths development). To investigate whether the impact of the

[18] This trial recruited a higher proportion of FSM pupils than is likely to be seen in the future implementation of 1stClass@Number 1. The model restricted to the subgroup of pupils not eligible for FSM was therefore added, providing additional context around the external validity of the primary analysis estimate. This addition to the subgroup analysis was not included in the Statistical Analysis Plan.

intervention was the same across the two cohorts we performed further subgroup analyses in which the sample was restricted to Cohort 1 and Cohort 2 pupils respectively.

For each potential moderator (variable defining a subgroup) an additional model with an interaction term was used to estimate the differential impact of the intervention in the subgroup. In the case of FSM, the differential effect for FSM-eligible pupils relative to non-FSM pupils was investigated by including an interaction term in the primary analysis model:

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \mathrm{intervention}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + \beta_4 \mathrm{FSM}_{ij} * \mathrm{intervention}_{ij} + b_j + \epsilon_{ij}$$

In this equation, β_4 quantifies the additional (or lesser) impact of the intervention for FSM-eligible pupils. Note that while this trial is powered to detect an intervention effect in the FSM subgroup, it is not designed to be powered to test whether there is a different effect for FSM pupils relative to non-FSM pupils (whether $\beta_4 = 0$). A similar interaction model was also calculated for each the other three subgroups described above: SEN pupils, pupils with SENT-R scores of 40 or below, and Cohort 1 pupils. For example, in the case of pupils with SEN the equation was:

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \mathrm{intervention}_{ij} + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + \beta_4 \mathrm{SEN}_{ij} + \beta_5 \mathrm{SEN}_{ij} * \mathrm{intervention}_{ij} + b_j + \epsilon_{ij}$$

where SEN_{ij} is an indicator for whether pupil i in school j has special educational needs.

Additional analyses and robustness checks

TA qualifications, experience, and confidence with maths

At the end of the intervention delivery period a survey was sent to all participating TAs, which included items on their qualifications, teaching experience, and their own confidence with maths. To investigate the extent to which they moderate the impact of 1st Class@Number 1 on maths attainment in the QRT for the four intervention pupils linked to each TA, four measures derived from these items were included in regression modelling:

(1) whether the TA achieved at least a grade C or equivalent in their maths GCSE (binary, yes or no);

(2) how many years of experience the TA has teaching maths to KS1 pupils (integer, modelled as a continuous variable);

(3) how many years of experience the TA has teaching KS1 pupils in any subject (integer, modelled as a continuous variable); and

(4) the TA's self-reported confidence in their own maths abilities; six Likert scale items relating to aspects of the TA's self-reported confidence with maths (see Appendix E) were summed to form a 6–24 scale, where higher values indicate more confidence (modelled as a continuous variable).

These were measures included in the selection guidance for the TA delivering 1st Class@Number 1 felt to be important moderators for the impact of 1st Class@Number 1 based on the intervention's theory of change. The fourth measure was bespoke to this evaluation and while it was not be validated extensively (for example through factor analysis), the Cronbach's Alpha of the scale was calculated as 0.93, indicating a high reliability. Each of these measures was included as a predictor in a linear multilevel model in which the individual observations were pupils, not TAs. These models were similar to the primary analysis except that only intervention pupils were included, and the intervention indicator was replaced by a variable representing measure (1), (2), (3), or (4) above. In the case of measure (1) the model was:

$$\mathrm{QRT}_{ij} = \beta_0 + \beta_1 \mathrm{TA\_GCSE\_math}_j + \beta_2 \mathrm{QRT\_BL}_{ij} + \beta_3 \mathrm{FSM}_{ij} + b_{0j} + \epsilon_{ij}$$

Here $\mathrm{TA\_GCSE\_math}_j$ is an indicator variable, taking the value 1 if the TA at school j achieved a grade C or above in GCSE maths. The models for measures (2), (3), and (4) followed similar lines, with $\mathrm{TA\_GCSE\_math}_j$ replaced by a variable for the appropriate measure. For measures (2) and (3), it was considered that the implied linear relationship between years of TA experience and pupil QRT score may not hold. We therefore plotted this relationship for both measures and planned to model the measures as appropriate if a clear non-linear relationship is observed (for example including a squared term for a quadratic relationship). Furthermore, as all four measures were likely to be correlated, they were also included simultaneously in a further model:

$$\mathrm{QRT_{ij}} = \beta_0 + \beta_1 \mathrm{TA\_GCSE\_math_j} + \beta_2 \mathrm{TA\_years\_math\_teaching_j} + \beta_3 \mathrm{TA\_years\_any\_teaching_j} + \beta_4 \mathrm{TA\_math\_confidence_j} + \beta_5 \mathrm{QRT\_BL_{ij}} + \beta_6 \mathrm{FSM_{ij}} + \mathrm{b_{0j}} + \epsilon_{ij}$$

In this model, $\mathrm{TA\_years\_math\_teaching_j}$, $\mathrm{TA\_years\_any\_teaching_j}$ and $\mathrm{TA\_math\_confidence_j}$ represent measures (2), (3), and (4) respectively for the TA at school j.

Sensitivity check—impact of online training

There were N = 16 schools where TAs attended only online rather than face to face training. Due to the inclusion of this online training group, we performed a sensitivity analysis that excluded all pupils from schools whose TAs had received online training, but otherwise followed the same model specification as the primary outcome:

$$\mathrm{QRT_{ij}} = \beta_0 + \beta_1 \mathrm{intervention_{ij}} + \beta_2 \mathrm{QRT\_BL_{ij}} + \beta_3 \mathrm{FSM_{ij}} + \mathrm{b_{0j}} + \epsilon_{ij}$$

This approach could also be viewed as a subgroup analysis, but interest lies more in establishing whether the primary analysis impact estimate is robust to the inclusion of the online training group (rather than quantifying the moderating effect of online training per se). A purely online training option is available to schools as part of the usual delivery of 1st Class@Number 1; EHU has noted that this format is becoming increasingly popular among schools.

Sensitivity check—allowing the impact of the intervention to vary between schools

For the primary analysis the choice was made to analyse this multisite trial using a mixed effects model with random school intercept only, as discussed in the Primary Analysis section. A limitation of this approach is that it assumes the impact of the intervention is the same for all schools (Feaster, Mikulich-Gilbertson and Brincks, 2011). If the assumption is not met, results may not be generalisable to schools outside the trial. To assess the sensitivity of the primary analysis to this assumption, the following model was calculated:

$$\mathrm{QRT_{ij}} = \beta_0 + \beta_1 \mathrm{intervention_{ij}} + \beta_2 \mathrm{QRT\_BL_{ij}} + \beta_3 \mathrm{FSM_{ij}} + \mathrm{b_{0j}} + \mathrm{b_{1j}} \mathrm{intervention_{ij}} + \epsilon_{ij}$$

Here $\mathrm{b_{0j}}$ and $\mathrm{b_{1j}}$ represent the school random intercept and random slope terms, respectively, which were allowed to be correlated in the model; $\mathrm{b_{1j}}$ was added to the primary analysis model, representing the effect of the intervention varying from school to school. This model was calculated both for all pupils and for the FSM subgroup with the $\mathrm{FSM_{ij}}$ variable removed in the second case. The distribution of school-specific intervention effects[19] was extracted and is plotted in the results section. The parameter ω, which is the proportion of school-level variance due to the impact of the intervention varying between schools, was calculated using the formula:

$$\omega = \frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{E}}^2}$$

In this formula, σ_{B}^2 is the between-school variation and σ_E^2 is variation due to the differing effect of the intervention across schools. Both are unconditional, so were obtained from a model with random intercept and slope but no fixed covariates.

Estimation of effect sizes

Impact estimates from all models described above were presented as an effect size:

$$\mathrm{ES} = \frac{\hat{\beta}_1}{\sqrt{\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{W}}^2}}$$

$\hat{\beta}_1$ is the estimated intervention or compliance coefficient from the appropriate model (conditional on covariates), while σ_B^2 and σ_{W}^2 are the between-school and within-school variance from an empty model. To obtain a 95% confidence interval

[19] That is, the overall fixed effect of the intervention plus the 'best linear unbiased predictor' of the random effect of the intervention for that school.

for this effect size, a confidence interval for $\hat{\beta}_1$ was first calculated using the profile likelihood,[20] where possible, as described by, for example, by Cole, Chu and Greenland (2013); otherwise by using the Wald method, which is $\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)$. The end points of this confidence interval were then divided by the denominator of the above effect size formula.

One exception to this approach occurred when calculating a confidence interval for the effect sizes estimated from the random slope multilevel models specified in the Additional Analysis section. In this case the variance of the effect size was calculated as:

$$Var(\mathrm{ES}) = \frac{1}{M\sigma_T^2}\left(\sigma_E^2 + \frac{2\sigma_W^2}{n}\right) + \frac{\mathrm{ES}^2}{4M(n-1)}$$

where σ_T^2 is the total outcome variation and σ_E^2 is variation due to the differing effect of the intervention across schools. M and n are the number of schools and number of pupils per school respectively. A confidence interval was then obtained by adding or subtracting $1.96 \times \sqrt{Var(\mathrm{ES})}$ from the effect size. A derivation for the above variance, as well as arguments for using it to calculate confidence intervals (instead of the simpler approach above) is given by Singh et al. (2022).[21]

Estimation of ICC

The ICC for the primary outcome model was calculated as the proportion of endpoint QRT score variance attributable to level 2 (between-school) variation:

$$\mathrm{ICC} = \frac{\sigma_\mathrm{B}^2}{\sigma_\mathrm{B}^2 + \sigma_\mathrm{W}^2}$$

Here, σ_B^2 and σ_W^2 are the between-school and within-school variation, which were extracted directly from the multilevel regressions fitted by the `lme4` R package (Bates et al., 2015). The ICC was calculated twice, once for the primary analysis model and once for an empty model (one with no covariates). These values are reported in the Results section, even though the ICC was not used in the sample size calculation for this evaluation.

$$\mathrm{ES} = \frac{\hat{\beta}_1}{\sqrt{\sigma_\mathrm{B}^2 + \sigma_\mathrm{W}^2}} \quad \hat{\beta}_1 Var(\mathrm{ES}) = \frac{1}{M\sigma_T^2}\left(\sigma_E^2 + \frac{2\sigma_W^2}{n}\right) + \frac{\mathrm{ES}^2}{4M(n-1)}$$

Methods—implementation and process evaluation

The implementation and process evaluation (IPE) was designed to complement the impact evaluation by exploring how 1stClass@Number 1 was implemented and how differences in implementation may affect intervention effectiveness. [22] It was also designed to build on previous evaluations of 1stClass@Number 1 by exploring elements of the logic model that were not fully explored previously (for example, perceived outcomes for pupils and TAs, parental engagement). Furthermore, because the trial used a pupil randomised design, we included a research question to explore how schools felt about this and how well they were able to implement it.

Research methods

The research team drew on a range of data collection methods to answer the IPE research questions (detailed below). An overview of the IPE methods is presented in Table 6.

[21] While the formula given is approximately correct, the number of pupils was slightly unbalanced amongst schools, which required a more complicated version of the formula for the effect size variance. This is also given by Singh et al. (2022).

[22] See IPE guidance for further details.

IPE RQ1: Who was selected to participate in the intervention and how were they chosen?

Reach

1.1 Which TAs were selected to deliver the intervention and how were they chosen? What was their previous experience of providing small group support and/or maths support to pupils?

1.2 How were pupils selected to participate in the intervention? To what extent were the selection criteria and guidance followed (including the focus on selecting FSM pupils) and how did schools feel about this?

IPE RQ2: To what extent was fidelity to (a) the intended training design and (b) the intended intervention delivery achieved?

Dosage

2.1 Did TAs and Link Teachers attend all the relevant training sessions (face to face or online)? If not, why not?

2.2 Did all pupils receive the intended intervention dosage (that is, the full 30 sessions in order—compliance measure 1)? If not, why not? Were there any differences between FSM and non-FSM pupils?

2.3 To what extent did pupils receive at least five out of six sessions of each topic, and at least 25 out of 30 sessions in total (compliance measure 2) and did this differ for FSM pupils?

Fidelity, adaptation, quality and responsiveness

2.4 Were the training and intervention sessions delivered in accordance with the intervention manual(s)? How variable was this between trainers and schools?

2.5 How did TAs and Link Teachers respond to and engage with the training, and did it provide them with the knowledge, skills, and confidence to deliver the intervention?

2.6 What were Link Teacher perceptions of TA delivery quality?

IPE RQ3: How effectively was the intervention implemented in schools and what supported or hindered successful implementation?

Barriers and facilitators and context

3.1 What were the barriers and facilitators to intervention implementation?

3.2 What contextual factors (for example, small group) supported or hindered successful intervention delivery?

3.3 To what extent did the support provided by EHU and Link Teachers facilitate successful intervention delivery?

IPE RQ4: What were the perceived benefits of the intervention for TAs and pupils?

Perceived outcomes (based on the logic model)

4.1 Did the training result in TAs having improved maths knowledge and/or confidence teaching maths and/or making any changes to their practice?

4.2 How did pupils feel they benefitted from the intervention (based on the logic model), for example, maths knowledge, maths confidence, maths self-efficacy, motivation towards maths and mathematical talk?

4.3 How did TAs feel pupils benefitted from the intervention and did they feel there were any differences between FSM and non-FSM pupils?

4.4 Were there any unexpected/unintended outcomes?

Moderators/responsiveness

4.5 Did TAs or Link Teachers feel that any pupils benefitted from the intervention more than others based on pupil characteristics including socioeconomic disadvantage? Why/why not?

4.6 How did pupils respond to and engage with the intervention? Why? What facilitated or hindered their engagement?

IPE RQ5: What was business as usual (BAU) and did this change during the intervention?

Usual practice

5.1 What was usual practice in intervention schools?

5.2 Did schools make any changes to their planned support during the trial and/or after learning the pupil allocation to intervention/control groups?

Monitoring of control group

5.3 Did schools provide any additional maths support (apart from 1stClass@Number 1) to either intervention or control pupils during the trial? If so, what maths interventions did intervention/control pupils receive and what was the dosage of these interventions?

5.4 Did TAs feel that they made any changes to their practice as a result of 1stClass@Number 1 which changed the way they worked in the classroom and/or with control pupils?

IPE RQ6: What are the facilitators, barriers, and features of delivering at scale and is the intervention sustainable?

Scale-up

6.1 What were the facilitators, barriers, and features of training and intervention delivery at scale?

6.2 What are EHU's plans for future training and delivery at scale?

Costs and sustainability

6.3 To what extent are schools able to continue resourcing and staffing the intervention going forward? Do they plan to continue using the intervention?

6.4 Would schools be willing and able to pay for the intervention without the subsidy?

6.5 To what extent is the intervention perceived to be cost-effective and offer good value for money?

IPE RQ7: How manageable was it for schools to facilitate the pupil-randomised design?

Context

7.1 How did schools feel about the randomisation of pupils to groups?

7.2 How well did the randomisation of pupils to groups work in practice? Did it create any challenges (for example, around managing the group)?

7.3 To what extent did schools adhere to the random allocation of pupils (that is, not switch children between groups)?

7.4 How well did schools manage the trial and minimise contamination between intervention and control pupils?

Table 6: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Achieved sample sizes	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Observations and reflective interviews	Structured observations and semi-structured reflective interviews	Trainer observations (one per trainer) TA observations (across topics and training groups)	N = 13 N = 10	Frequency counts; deductive coding; thematic analysis	RQ2, RQ3	Fidelity and adaptations, quality, responsiveness, facilitators and barriers
Registers	Training attendance registers Pupil intervention attendance registers Pupil maths intervention records	Registers expected from all training groups Intervention registers expected for all intervention pupils Maths intervention records expected for all pupils	N = 17 N = 838 N = 1561	Frequency counts; descriptive statistics	RQ2; RQ5	Compliance and dosage, monitoring of control group, usual practice
Surveys	Online surveys (endpoint)	All Link Teachers and TAs (excluding those from schools that did not deliver the intervention) invited to respond	Link teacher N = 174 TA N = 187	Descriptive statistics; sub-group analysis	RQ1, RQ2, RQ4, RQ6	Reach, fidelity and adaptations, responsiveness, perceived outcomes, moderators
Case study interviews	Semi-structured interviews	12 Link Teachers and TAs across 12 schools	Link teacher N = 12 TA N = 12	Deductive coding and inductive coding; thematic analysis	RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7	Reach, usual practice fidelity and adaptations, responsiveness, quality, facilitators and barriers, perceived outcomes, scale up, content
Pupil focus groups	Focus groups involving creative activities	Intervention pupils across 12 case study schools	N = 43	Deductive coding and inductive coding; thematic analysis	RQ4	Responsiveness, perceived outcomes

Training observations and reflective interviews

Training observations were conducted across 13 of the 17 training groups. Each trainer was observed once, with the exception of one trainer who delivered both an in-person and online group and was therefore observed twice. The remaining four training groups were not observed because the trainers who delivered these groups led multiple groups but were only observed once each. Observations were spaced throughout the training period and across the two cohorts (eight observations were conducted in Cohort 1, five in Cohort 2), which allowed the research team to observe each of the four training sessions covering the intervention content more than once. This meant the research team could check consistency of delivery across the trainers and regions and between cohorts. Observations of the fifth training session were not conducted as this session took place after the end of delivery to pupils and did not contain any of the pupil intervention content (the fifth training session is for TA and Link Teachers' professional reflections).

An observation schedule was designed by NFER and populated by a member of the research team at each of the training sessions they attended. These observations focused on how the training was delivered by the 12 trainers, any differences in quality, engagement, style of delivery, or adaptations to content, and staff engagement with and response

to the training session. Following each observation, trainers participated in a reflective interview which captured their experiences of delivering the training and their perceptions of engagement at the sessions. TAs and Link Teachers were able share their views about training in surveys and during interviews if they took part in case studies (see below).

Lesson observations and reflective interviews

Lesson observations of TAs delivering the intervention were conducted in ten schools. These schools were initially sampled based on deprivation, cohort size, and location and were contacted ahead of the training session. However, observation dates were more often scheduled by the researcher in coordination with the TA after the training session as it was easier to discuss requirements and confirm a suitable time in-person rather than via email. Of the ten observations, seven took place in Cohort 1 schools and three in Cohort 2 schools, including an observation of one TA who received online training. The research team initially planned to observe two lessons from each of the five topics, however, due to TA illness, which led to a postponed visit, topic four was observed only once, while three lessons from topic five were observed.

An observation schedule was designed and populated by a member of the NFER research team at each of the lesson observations. Observations focused on the setting in which the intervention was delivered, resources used, the extent to which the full session content was covered, correct use of mathematical language, appropriate adaptations to pupils' needs, pupil participation in the session, and perceived outcomes. Following each observation, TAs participated in a reflective interview which explored their experiences of delivering the session and their perceptions of pupil engagement, as well as their views about early outcomes they had observed.

Schools selected to participate in a lesson observation were not selected to be a case study school to minimise burden on schools and ensure a wider range of schools were represented in the data collection activities.

Attendance registers and intervention records

Link teacher and TA attendance was recorded at each training session to monitor training dosage and compliance. EHU ordinarily collect attendance data so usual practice systems were in place and this data was shared with NFER.

Outside of the trial, EHU typically provide TAs with a paper register to record pupil attendance at intervention sessions. At the first training session, TAs were provided with a paper register, adapted by NFER for the purpose of the trial. Link Teachers and TAs were asked to upload this attendance data into an online register created by NFER in Excel so it could be shared with NFER at the end of the trial. In addition to the attendance register, NFER designed an intervention record which Link Teachers completed for all eight of the pupils (intervention and control) involved in the trial. This was used to understand any other maths interventions pupils had participated in over the course of the trial. This data was also collated in an Excel template and shared with NFER at the end of the trial.

Surveys

The Link Teacher and TA surveys were administered online at endpoint. Surveys of staff in Cohort 1 were administered in the spring term of 2024 (4 March to 19 April) and surveys of Cohort 2 staff were administered in the summer term (20 May to 14 June). Survey timing was carefully planned to ensure that surveys were open at the end of each delivery window and remined available until after TAs had delivered all the intervention sessions and had attended the final training session. Overall, Link Teachers from 174 and TAs from 187 of the 217 schools that remained in the trial completed the surveys (that is 80% and 86% response rates respectively). [23] By cohort: Link Teachers from 122 and TAs from 125 of the 153 Cohort 1 schools responded to the surveys; in Cohort 2, Link Teachers from 52 and TAs from 62 of the 64 schools responded. These surveys explored Link Teachers' and TAs' views about the received training, experiences of implementing the intervention, perceived pupil and parent/carer engagement, and perceived outcomes. In addition, TAs were asked to respond to questions about their experience as a TA and their confidence in their own

[23] Surveys were not sent to Link Teachers and TAs in schools which withdrew from intervention delivery because their reasons for withdrawing were capacity related. Additionally, their early withdrawal would have limited their ability to respond to survey questions related to the intervention training and perceived outcomes.

maths skills and abilities. Link Teachers were also asked about how TAs were selected to deliver the intervention, their plans for delivering 1st Class@Number 1 in the future, and their views about value for money.

School case studies

Case studies were conducted with 12 schools—eight from Cohort 1 and four from Cohort 2. Case study schools were located across 11 of the 17 geographical regions where the trial was delivered, with one case study school selected from the online training group. Case studies of Cohort 1 schools took place between March and June 2024 while Cohort 2 case studies took place between June and July 2024. School case study visits took place after the school had completed its endpoint testing. The research team aimed to conduct these visits within one to two weeks following the endpoint testing to ensure the intervention was still fresh in pupils' minds. Only one Cohort 1 case study took place over two weeks after endpoint testing due to capacity challenges associated with SATs. Case study schools were purposively sampled to ensure schools:

- reflected a range of deprivation levels (low, medium and high proportions of FSM);
- differed in cohort size;
- were from different training groups and regions across the country; and
- were a mix of urban and rural schools.

As noted, schools invited to be a case study were different to those invited to participate in a lesson observation to reduce burden on schools. Initial sampling took place in autumn 2023 with further sampling taking place in spring and summer 2024 to identify additional schools to participate in the case studies. In total, 27 schools were contacted to participate in case studies. Despite additional sampling taking place to replace schools initially selected that could not participate due to capacity challenges, the research team still achieved a final sample which represented schools with a range of deprivation levels, cohort sizes, and training regions.

In each of the 12 case study schools, the Link Teacher and TA participated in separate interviews tailored for their role. Interviews explored:

- reasons for participating in the trial;
- experiences of the training;
- implementing the intervention in school and associated facilitators and barriers;
- perceived outcomes for staff and pupils;
- views on parent/carer engagement with the intervention;
- views on value for money; and
- plans for delivering the intervention in the future, including factors which would support or hinder sustainability (Link Teachers only).

In total, 43 (of the 48) intervention pupils across the 12 case study schools participated in pupil focus groups. These focus groups were designed to build on previous evaluations by exploring pupil perspectives on the intervention. The research team supplied Link Teachers with parent/carer information and consent forms so that opt-in parent/carer consent for their child to participate in the group could be gained ahead of the case study visits. The pupil focus groups included three creative activities suitable for young children to support pupils' engagement with the researchers' questions. The first activity required pupils to identify what they had liked, disliked, or would change about the intervention (responsiveness) which they or (where needed) the researcher wrote on Post-it notes. The second activity involved creating a learning tree to display what pupils felt they had learnt or improved at from participating in the intervention (perceived outcomes). In the final activity, pupils designed a postcard to show how they had engaged with the special delivery activities at home or at school (responsiveness) to understand if/how these activities were used.

Delivery team member interview

In summer 2024, the research team conducted an online interview with a member of the 1stClass@Number 1 delivery team. This interview helped the research team to understand the delivery team's views on the extent to which the training had been delivered with fidelity to the intended approach and their perceptions of delivery quality. The interview also covered their experiences of delivering the intervention at scale and views on the sustainability of future scale-up, including facilitators and barriers to scaling up.

Data collection instruments

With the exception of the attendance registers, the IPE research team developed all of the data collection instruments used in the IPE: intervention records, schedules for the training and lesson observations, interview schedules for the reflective interviews, case study interviews with Link Teachers, TAs, and with the delivery team, materials to conduct the pupil focus group activities, and the endpoint Link Teacher and TA surveys. The research team collected all data from the surveys and interviews. Training attendance registers and pupil attendance registers were based on the registers developed and used by EHU and adapted by the IPE team for the purposes of the trial. EHU collected data on Link Teacher and TAs attendance at the 1stClass@Number 1 training sessions and shared this with NFER via a secure data sharing portal. Link Teachers and TAs collected data on intervention pupils' attendance at the 1stClass@Number 1 sessions and on any other maths interventions which each intervention and control group pupil had received over the course of the trial. These were also shared with NFER via the secure portal.

Rationale for the data collection methods utilised

NFER researchers chose the range of data collection methods outlined above to achieve both breadth and depth for the IPE. The logic model was used to help prioritise data collection to focus on the key features of the intervention and the assumptions underpinning it. The registers provided an efficient way of collecting Link Teacher and TA training attendance as well as pupil attendance at the intervention. Both registers were designed to then feed into the compliance measures for the trial. The maths intervention record, for intervention and control pupils, provided an efficient way of understanding any other maths interventions received over the course of the trial (akin to BaU/what else pupils were receiving), including a check of whether any control pupils had experienced 1stClass@Number 1 (contamination). The surveys provided an efficient way of measuring experiences of the training and implementation of the intervention across a large number of schools. They also provided the opportunity for TAs to report perceived professional outcomes. The TA survey also included questions about their mathematical knowledge and experience delivering small group interventions, which were used in the impact evaluation. The qualitative data collection activities with trainers, school staff, pupils, and the delivery team provided more in-depth insights into the implementation, perceived outcomes, and sustainability of 1stClass@Number 1. Since the children involved in this trial are quite young (six to seven years old), creative activities were used during the pupil focus groups to support their engagement with the researcher's questions.

Analysis

Qualitative data—observations, interviews, and pupil focus groups

The research team took detailed notes during observations. Interviews and pupil focus groups were audio recorded then written up as verbatim transcripts. The research team used the qualitative data software MAXQDA to code the qualitative data. A coding frame was developed based on the theory of change, IPE dimensions, and research questions to cover the key areas of interest and was used to code all data sources then consider the diversity, frequency, and strength of views about the programme.

Transcripts and observations notes were labelled according to participant type, school, training region, and cohort before being uploaded into MAXQDA as this would allow the research team to compare perspectives across groups. High level deductive coding was used to collate evidence from across the different data sources under each dimension within the coding framework. A round of inductive coding was then used to identify key themes within each of these dimensions. In addition to coding transcripts from pupil focus groups, the research team thematically analysed the creative outputs from these groups by grouping pupils' written responses by theme.

Quantitative data—surveys and attendance registers

Quantitative data from the surveys and attendance registers was primarily analysed using descriptive statistics. Descriptive statistics included percentage response rates to different response options and, where applicable, median, averages, and standard deviations. Cross group analysis, by cohort and whether training was received in-person or online, was also conducted for relevant survey questions.

In addition to descriptive statistics of survey responses, composite scores of TAs' maths knowledge and confidence were created to explore the potential of these to moderate the intervention outcomes. The composite measure was included in the regression models used for the impact evaluation to undertake exploratory analysis of these intervention moderators.

Triangulation of qualitative and quantitative data

The research team triangulated the IPE qualitative and quantitative data sources in order to cross-validate the results and support the interpretation of the findings. This was achieved through designing the IPE instruments simultaneously and creating an analysis framework which mapped out how the themes generated from the qualitative data analysis and questions from the quantitative surveys linked to each other and to the IPE dimensions included in the report. In addition, responses to TA survey questions related to maths qualifications, teaching experience, and maths confidence were included in the regression models to explore whether these moderated intervention outcomes (see TA Qualifications, Experience and Confidence with Maths for more details about this analysis).

Costs

A full cost evaluation of the 1stClass@Number 1 programme was previously undertaken as part of the previous effectiveness trial (Nunes et al., 2018). There have been no changes to the programme design or implementation that are likely to have affected the time or resource costs associated with the intervention since this cost evaluation was undertaken. Therefore, to maximise efficiency, it was decided that this evaluation would not undertake another full cost evaluation. Instead, the research team agreed with the EEF to update the costs from the previous evaluation, adjusting for inflation and incorporating actual costs where available (for example, the current price of the SENT-R for schools).

Timeline

Table 7: Timeline

Dates	Activity—Cohort 1	Activity—Cohort 2	Organisation responsible
Dec 2022–Feb 2023	Project set-up. Complete project set-up, finalise recruitment documents.		NFER with EHU
Mar–Sep 2023	Recruitment and pupil data collection. Recruitment of Cohort 1 and Cohort 2 schools by EHU: recruitment of Cohort 1 schools to be completed by end of summer term, recruitment of Cohort 2 schools to continue until end of third week of September. EHU share school data for Cohort 1 and 2 schools with NFER; NFER contact schools to collect pupil data for nominated pupils and TA and Link Teacher data. NFER prioritise collection of pupil data from Cohort 1 schools before end of summer term. Collection of pupil data from Cohort 2 schools ongoing, to be completed in the second week of October.		EHU (recruitment), NFER (school/pupil/TA data collection)
	Study Protocol Submit draft protocol, revise based on feedback and publish study protocol.		NFER, EEF
Jul–Aug 2023	Preparation of assessment materials and administration guidance for Cohort 1 schools. NFER prepare Sandwell Early Numeracy Test – Revised (SENT-R) and baseline		NFER

	QRT test booklets for nominated pupils in Cohort 1, develop guidance for teachers to administer SENT-R and QRT, develop guidance for processing SENT-R scores.		
	Start design of IPE instruments.		NFER
Sep 2023	SENT-R screening and baseline QRT and selection of pupils to trial in Cohort 1 schools. SENT-R and baseline QRT administered by Cohort 1 schools; test booklets sent to NFER for marking/processing; NFER send SENT-R score report to schools recommending pupils to be selected to the trial as per prioritisation strategy; schools confirm final eight pupils for trial.	Complete pupil data collection from Cohort 2 schools.	NFER; schools administer baseline
	Complete design of IPE instruments.		NFER
Oct 2023	Randomisation of pupils in Cohort 1 schools.	Preparation of assessment materials for Cohort 2 schools. NFER prepare SENT-R and baseline QRT test booklets for nominated pupils in Cohort 2.	NFER
	Notify Cohort 1 schools of study arm to which each pupil is assigned.		NFER
	Complete marking and QA of baseline QRT data from Cohort 1 schools.		NFER
	Submit draft Statistical Analysis Plan (SAP).		NFER
Nov–Dec 2023	See below.	SENT-R screening and baseline QRT and selection of pupils to trial in Cohort 2 schools. SENT-R and baseline QRT administered in Cohort 2 schools; test booklets sent to NFER for marking/processing; NFER send SENT-R score report to schools recommending pupils to be selected to the trial as per prioritisation strategy; schools confirm final eight pupils for trial.	NFER; schools administer baseline
		Randomisation of selected pupils in Cohort 2 schools.	NFER
Nov 2023–Mar 2024	Training and intervention delivery and data collection for Cohort 1 schools 1stClass@Number 1 training and intervention delivery for Cohort 1 schools.	See below.	EHU, schools
	IPE data collection in Cohort 1 schools—trainer and TA observations and reflective interviews.		NFER
	Submit NPD data application (Jan 2024).		NFER
	QRT endpoint testing in Cohort 1 schools by NFER test administrators; TA/ LT surveys (Mar 2024)	Complete marking and QA of baseline QRT data from Cohort 2 schools.	NFER; NFER test administrators
	Obtain pupil attendance/other intervention records from Cohort 1 schools (Mar 2024).		NFER

	Case study visits for Cohort 1 schools (Mar 2024).		NFER
Jan–Jun 2024	See above and below.	Training and intervention delivery and data collection for Cohort 2 schools. Training and intervention delivery for Cohort 2 schools.	EHU, schools
		IPE data collection in Cohort 2 schools—trainer and TA observations and reflective interviews.	NFER
		QRT endpoint testing in Cohort 2 schools by NFER test administrators; TA/ LT surveys (May/Jun 2024).	NFER; NFER test administrators
	Obtain training attendance data for both cohorts from EHU (Jun ’24).		NFER, EHU
		Obtain pupil attendance/other intervention records from Cohort 2 schools (Jun 2024).	NFER
		Case study visits for Cohort 2 schools (May/Jun 2024).	NFER
	Publish Statistical Analysis Plan (SAP).		NFER, EEF
Apr–May 2024	Complete marking and QA of endpoint QRT data from Cohort 1 schools.		NFER
Jun–Jul 2024		Complete marking and QA of endpoint QRT data from Cohort 2 schools.	NFER
Jul–Aug 2024	QA and analysis of IPE data.		NFER
Sep–Nov 2024	Access data on NPD, complete primary and additional analyses.		NFER
Jan 2025	Submit first draft of report.		NFER
Jun/Jul 2025	Final report published.		NFER, EEF
Sep 2025	Submit data to EEF archive.		NFER

Impact evaluation results

Participant flow including losses and exclusions

The participant flow diagram for this evaluation is shown below (Figure 2). In the recruitment phase, 325 schools were initially approached by EHU of which 248 agreed to take part in the trial; 22 of the 248 schools dropped out before randomisation, typically due to lack of staffing resource or the training location not being suitable. From the remaining 226 schools, 2,833 pupils were nominated as suitable for receiving 1st Class@Number 1, an average of 12.5 per school. After screening using the SENT-R, 1,797 pupils were then selected for randomisation, eight per school at most schools (as required by the evaluation design). Most pupils were selected using the criteria provided by NFER (N=1,725) although in some cases schools replaced the pupils selected by NFER with their own choices from the screened list and following instructions provided by NFER (N = 72: see IPE RQ1, Selection of Pupils section for detail). The pupils selected using NFER's criteria were prioritised according to their FSM status and SENT-R score, with pupils at a school from lower priority groups selected only if there were no pupils remaining from higher priority groups. From highest to lowest priority, the priority groups were:

(1) FSM pupils with SENT-R scores of 49 or below (N = 1,142).

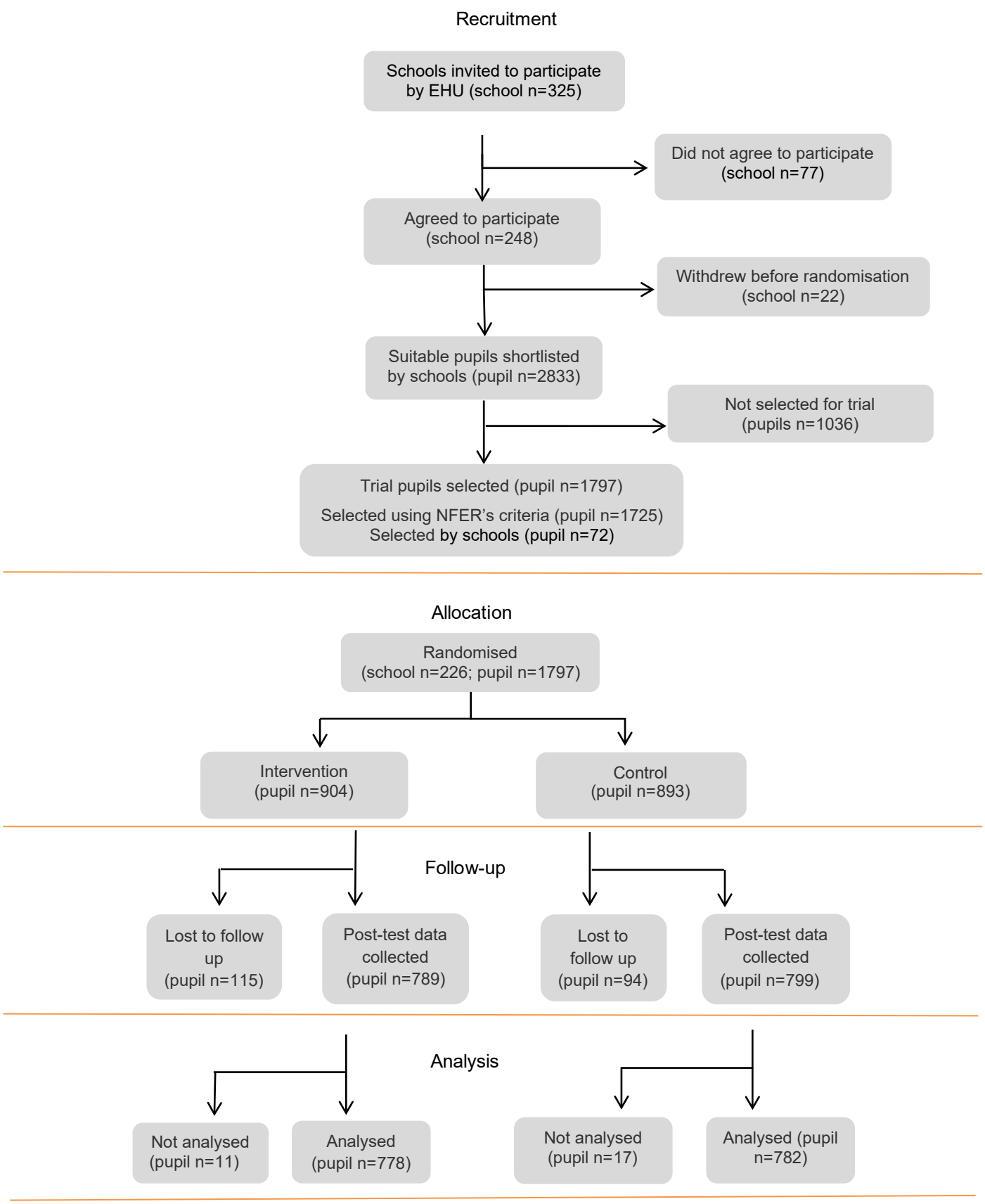
(2) Non-FSM pupils with SENT-R scores of 40 or below (N = 546).

(3) Non-FSM pupils with SENT-R scores of 41-49 (N = 94).

(4) Any pupils with SENT-R scores above 49 (N = 15).

Of the 1,797 pupils randomised, 904 were allocated to the intervention and 893 to the control. Between randomisation and analysis, 209 pupils did not provide an endpoint QRT score (intervention N = 115, control N = 94) and so were lost to follow-up. This was largely due to pupils being absent on the day of the endpoint QRT and pupils leaving the school (see the missing data results for a complete breakdown of the reasons for loss to follow-up). A further 28 pupils (intervention N = 11, control N = 17) were followed up but were not included in the complete case primary analysis due to a missing baseline QRT score or missing FSM status on the NPD. The final primary analysis sample, therefore, consisted of 1,560 pupils (intervention N = 778, control N = 782), 86.8% of those randomised.

Figure 2: Participant flow diagram



The MDES at the protocol, randomisation, and analysis stages is shown in Table 8. A 10% school attrition rate was predicted between recruitment and randomisation: this quantity has already been deducted from the number of schools given at the protocol and analysis stages shown below. The protocol and randomisation calculations also assumed a 15% pupil-level attrition rate. At the analysis stage, the parameters in the MDES calculation were generally more favourable than anticipated in the protocol: attrition rates were lower, the pre-post test correlation was higher, and the proportion of FSM pupils was higher. As a result, the MDES among all pupils reduced from 0.14 at the protocol stage to 0.13 at the analysis stage and among FSM pupils reduced from 0.2 to 0.16.

Table 8: Minimum detectable effect size at different stages

		Protocol		Randomisation		Analysis	
		Overall	**FSM**	**Overall**	**FSM**	**Overall**	**FSM**
MDES		0.14	0.2	0.13	0.16	0.13	0.16
Pre-test/post-test correlations	Level 1 (pupil)	0.3	0.3	0.3	0.3	0.44	0.39
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Number of schools		211	211	226	226	224	224
Average number of pupils per school	Intervention	4	2	4	2.6	3.5	2.2
	Control	4	2	4	2.5	3.5	2.2
	Total:	8	424	8	5.1	7.0	4.4
Total pupils at all schools	Intervention	844	422	904	577	778	494
	Control	844	422	893	573	782	499
	Total:	1688	844	1797	1150	1560	993

Attrition

The pupil attrition rate between randomisation and the primary analysis was moderate at 13.2% (Table 9). The requirement for pupils to attend endpoint testing on the QRT contributed to this attrition, primarily due to pupils being absent on the day of the test and also on any mop-up dates (see Primary Outcome section and see missing data analysis). This moderate level of attrition was not unexpected: a 15% rate had been anticipated given the post-pandemic context and ongoing challenges with school absences, even in primary schools. The attrition rate was slightly higher for

[24] It was assumed at the protocol stage that 50% of pupils randomised across all schools would be eligible for FSM, so the average number per school would be four. However, the sample size calculation does not require exactly the same number of FSM-eligible pupils at each school: only the overall proportion matters.

pupils randomised to the intervention group (13.9%) compared to those in the control group (12.4%). Further discussion on attrition is provided in the Missing Data section.

Table 9: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	904	893	1797
	Analysed	778	782	1560
Pupil attrition (from randomisation to analysis)	Number	126	111	237
	Percentage	13.9	12.4	13.2

Pupil and school characteristics

The baseline characteristics of schools in the trial compared with all primary schools in the country are shown in Table 10. Most noticeable is the high proportion of FSM pupils in trial schools: almost three quarters (73%) of trial schools were in the highest 40% of primary schools nationally. This is due to the trial's recruitment strategy, which targeted schools with a high proportion of FSM pupils including from within EIAs (see Participant Selection). The distribution of other trial school characteristics also differed compared to primary schools nationally. The size of the difference was small in most cases but was more pronounced with regard to location: 91% of trial schools were located in urban areas, compared to 70% nationally. The trial's recruitment strategy in conjunction with the focus on EIA regions might have contributed to the high proportion of schools from urban areas. Conversely, only 8% of trial schools (compared to 30% nationally) were from rural areas, possibly due to the requirement that each school nominate at least eight suitable pupils for the trial. The proportion of pupils reaching the expected standard in KS2 maths was lower, on average, at trial schools (mean of 70.4 versus 73.1) and varied less (standard deviation of 13.4 versus 16.6). These further differences were probably a by-product of targeting schools with a high proportion of FSM pupils—urban schools have a higher proportion of FSM pupils, on average, than rural schools.

Table 10: Baseline characteristics of trial schools compared with all primary schools in England

	Trial schools		Primary schools in England *	
Categorical variable	n/N	%	n/N	%
Establishment type				
Academies	95/226	42	6358/15413	41
Local authority-maintained schools	125/226	55	8039/15413	52
Free Schools	3/226	1	268/15413	2
Other school type	0/226	0	745/15413	5
Missing	3/226	1	3/15413	<1
Overall Ofsted rating				
Inadequate	4/226	2	202/15413	1
Requires Improvement	26/226	12	1051/15413	7
Good	177/226	78	11412/15413	74
Outstanding	16/226	7	1824/15413	12
Missing	3/226	1	924/15413	6
Urban or rural location				
Urban	205/226	91	10723/15413	70

Rural	18/226	8	4648/15413	30
Missing	3/226	1	42/15413	<1
% FSM pupils' national quintile **				
Lowest 20%	7/226	3	2965/15413	19
2nd lowest 20%	21/226	9	2963/15413	19
Middle 20%	30/226	13	2965/15413	19
2nd highest 20%	74/226	33	2965/15413	19
Highest 20%	91/226	40	2960/15413	19
Missing	3/226	1	595/15413	4
Continuous variable	n (missing)	Mean (SD)	n (missing)	Mean (SD)
% of pupils meeting the expected standard in maths at KS2	197 (29)	70.4 (13.4)	13428 (1985)	73.1 (16.6)

* Including middle deemed primary schools, but not all-through or infant schools.
** These quintiles were constructed using the proportion of pupils that had ever been eligible for FSM at every primary school in England, which is why there are almost the same number of these schools in each quintile on the right side of the table.
Percentages may not sum to 100 due to rounding.

At the pupil level, baseline characteristics were well balanced between the intervention and control groups, with only minimal differences across all variables (Table 11). The distribution of QRT scores was similar between the groups, in terms of means, standard deviations, and the overall shape of their distributions (Figure 3, the equivalent figure for FSM pupils is in Appendix F). Effect sizes for the baseline difference in scores (intervention minus control) were very small for both the QRT (0.03) and SENT-R (-0.02). Overall, these minor imbalances are not considered to pose a threat to the internal validity of the trial.

Consistent with the pattern observed at the school level, the proportion of FSM pupils in the sample was exceptionally high in both the intervention and control groups (64% overall). This is significantly higher than the FSM proportion amongst Year 2 pupils nationally (22%)[25] and likely exceeds the proportion in the 'target' population[26]—Year 2 pupils receiving 1stClass@Number 1 outside this trial. Because the trial sample is not representative of the usual population in terms of FSM status, questions arise regarding the external validity of the primary analysis result. However, the fact that the primary analysis result closely aligns with the findings for the FSM subgroup provides some reassurance, as discussed in the Subgroup Analysis section below.

Table 11: Baseline characteristics of pupils as randomised

	Intervention group		**Control group**	
Categorical variable	n / N	%	n / N	%
Eligible for FSM				
Yes	574 / 904	64	576 / 893	65
No	323 / 904	36	314 / 893	35
Missing	7 / 904	<1	3 / 893	<1
Gender				
Male	430 / 904	48	417 / 893	47
Female	467 / 904	52	473 / 893	53
		<1		<1

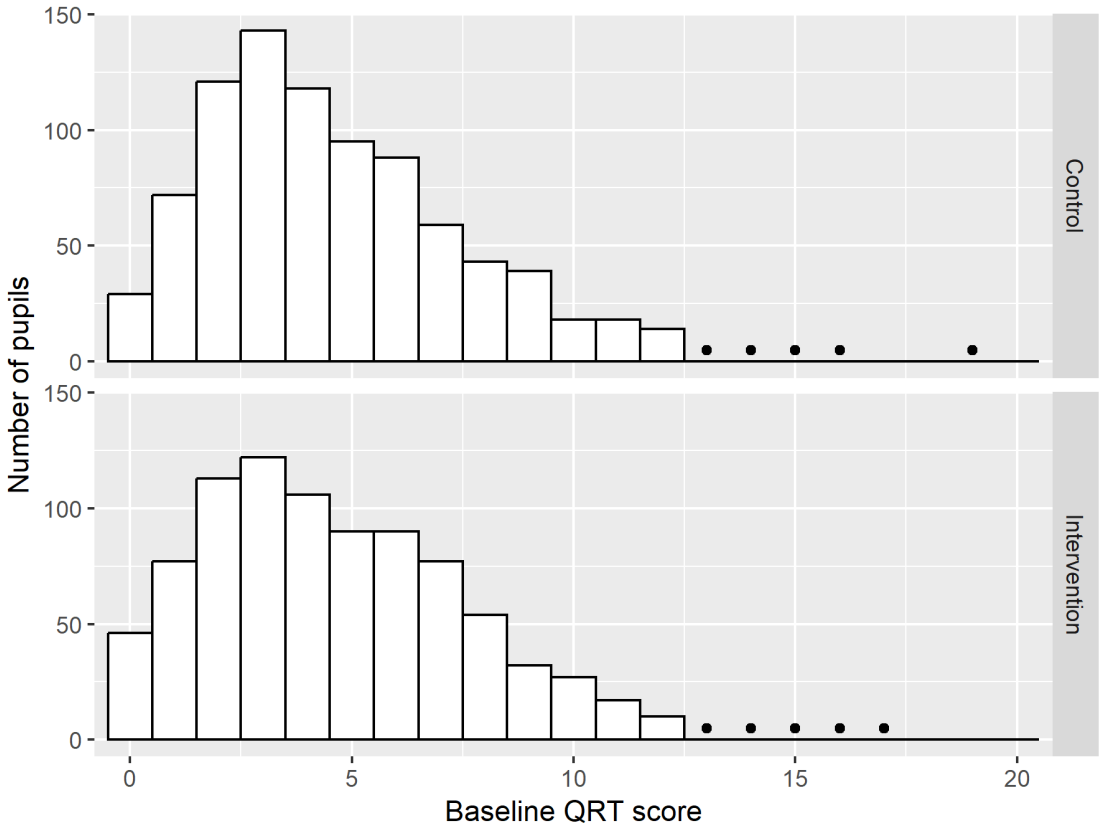
[25] Figure obtained at https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics

[26] Given the correlation between FSM status and experiencing moderate maths difficulties, the FSM proportion amongst pupils selected for 1stClass@Number 1 (outside this trial) should be higher than the national FSM proportion.

Missing	7 / 904		3 / 893		
Has SEN Yes No Missing	322 / 904 575 / 904 7 / 904	36 64 <1	327 / 893 563 / 893 3 / 893	37 63 <1	
Continuous variable	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect size (95% CI) *
Baseline QRT score	884 (20)	4.8 (3.2)	873 (20)	4.7 (3.1)	0.03 (-0.05, 0.11)
SENT-R score **	902 (2)	30.9 (9.5)	892 (1)	31.0 (9.2)	-0.02 (-0.09, 0.05)

* The average difference between continuous variables at baseline (intervention minus control) was calculated using a multilevel linear regression in which the outcome was the baseline variable, and the only predictor was randomisation status. The estimate from the regression and its confidence interval were then converted into an effect size.
** Having a baseline QRT score recorded was not a requirement for randomisation, but a SENT-R score was required. The three pupils with missing SENT-R scores did originally provide this information but later withdrew permission to have their data processed.
Percentages may not sum to 100 due to rounding.

Figure 3: Distribution of baseline QRT scores in the control and intervention groups



Black dots on the histogram represent pupil counts greater than zero but less than ten. These have been supressed to ensure individual pupils cannot be identified from the figures in this report.

Outcomes and analysis

Primary analysis

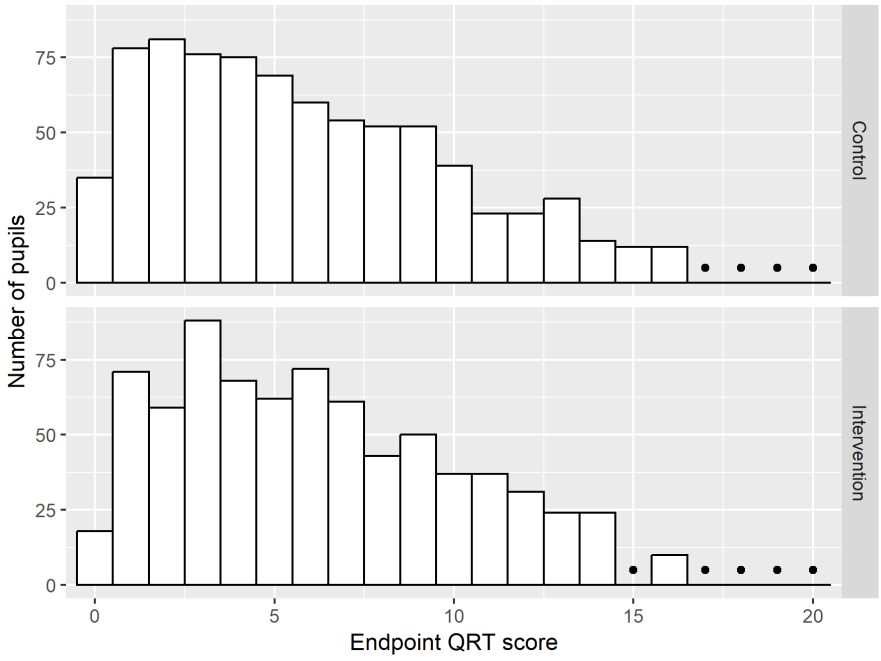
Based on the primary analysis model, pupils who received 1stClass@Number 1 scored, on average, 0.52 points higher on the endpoint QRT than control pupils (95% CI: 0.22 to 0.82), corresponding to an effect size of 0.12 (95% CI: 0.05 to 0.19, see Table 12). This is the best estimate of the intervention's impact but a range of small to moderate positive effect sizes are also supported by the data to a lesser extent. Notably, zero impact is not supported by the data, so it is extremely likely that 1stClass@Number 1 improves maths attainment as measured by the QRT for pupils who received the intervention compared to those who did not.

Figure 4 illustrates the distribution of endpoint QRT scores for both the control and intervention groups. There is a general improvement in scores from baseline in both groups but the most notable difference in the intervention group was a reduction in the number of pupils scoring zero at endpoint (N = 18) compared to baseline (N = 46), a change not observed in the control group. The distribution of endpoint QRT scores for FSM pupils in the control and intervention groups was similar to that of all pupils (Appendix F).

Table 12: Primary analysis results

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Effect size (95% CI)**	**p-value**
QRT score	789 (113)	6.64 (6.33, 6.95)	799 (107)	6.11 (5.80, 6.41)	1560 (778; 782)	0.12 (0.05, 0.19)	<0.001

Figure 4: Distribution of endpoint QRT scores in the control and intervention groups



Black dots on the histogram represent pupil counts greater than zero but less than ten. These have been supressed to ensure individual pupils cannot be identified from the figures in this report (similar distributions of QRT scores for FSM-eligible pupils can be seen in Appendix F).

Subgroup analyses

Since one of the main goals of this evaluation was to estimate the impact of 1stClass@Number 1 for the subgroup of FSM eligible pupils, this is the most important result from the subgroup analyses (Table 13). The effect size in the FSM subgroup was 0.11 (95% CI: 0.02 to 0.20), similar to the effect size of 0.12 seen in the primary analysis. The range of values supported by the data included small and moderate positive effects (but not zero), which is again similar to the primary analysis. As well as being of interest in itself, this result provides some reassurance that the primary analysis result is externally valid given the disproportionately high number of FSM pupils in the trial sample. The effect size in the non-FSM subgroup of 0.14 (95% CI: 0.02 to 0.25) was also recorded for context: an attempt to reweight the data to reflect a population with a lower proportion of FSM pupils would produce an effect size below this but above that seen in the primary analysis. The effect sizes seen in other pupil subgroups were all small and positive. Among these, the effect size for SEN pupils was the smallest (0.06, 95% CI: -0.07 to 0.18) and deviated the most from the primary analysis, although this difference may be due to chance, as described in the next paragraph. For pupils with a SENT-R score of 40 or below the effect size was 0.14 (95% CI: 0.06 to 0.21), slightly larger than in the primary analysis. A larger effect size for this subgroup might have been expected given that 1stClass@Number 1 is intended for pupils with moderate maths difficulties, although again the observed difference could be due to chance. The effect size seen amongst Cohort 1 pupils (0.14, 95% CI: 0.06 to 0.22) was somewhat larger than for Cohort 2 pupils (0.08, 95% CI: -0.06 to 0.21).

The differential impact of the intervention for pupils in each subgroup (for example, for FSM pupils compared with non-FSM pupils) is shown in Table 14, columns three and four. In each model the confidence interval for the interaction term contains zero and p-values are not close to zero. This means that there is insufficient evidence to conclude that each variable (for example, FSM status) moderates the impact of the intervention, based on the data in this trial. However, these interaction models are likely to be underpowered to detect small or moderate differential impacts.

By combining the appropriate terms from the interaction models, the effect size within a particular subgroup can also be estimated (Table 14, columns five and six), as required by the EEF Statistical Analysis Guidance (EEF, 2022). This is an alternative approach to obtaining the subgroup effect sizes seen in Table 14. The effect sizes obtained from the interaction models are very similar to those seen in Table 14. For example, the effect size for FSM is the same (0.11) in both cases, to two decimal places.

Table 13: Results within each pupil subgroup

	Unadjusted means				Effect size		
	Intervention group		Control group				
Subgroup	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Effect size (95% CI)**	**p-value**
FSM pupils	500 (74)	6.42 (6.04, 6.8)	512 (67)	5.88 (5.5, 6.26)	993 (494; 499)	0.11 (0.02, 0.20)	0.015
Non-FSM pupils	289 (34)	7.01 (6.47, 7.56)	287 (35)	6.52 (6.02, 7.01)	567 (284; 283)	0.14 (0.02, 0.25)	0.019
SEN pupils	274 (48)	5.37 (4.85, 5.89)	296 (46)	5.19 (4.68, 5.7)	558 (269; 289)	0.06 (-0.07, 0.18)	0.385
Pupils with SENT-R score ≤40	693 (98)	6.20 (5.88, 6.53)	695 (91)	5.55 (5.25, 5.86)	1363 (684; 679)	0.14 (0.06, 0.21)	<0.001
Cohort 1 pupils	546 (77)	6.74 (6.36, 7.13)	553 (76)	6.07 (5.7, 6.44)	1082 (538; 544)	0.14 (0.06, 0.22)	<0.001
Cohort 2 pupils	243 (36)	6.40 (5.87, 6.93)	246 (31)	6.19 (5.67, 6.7)	478 (240; 238)	0.08 (-0.06, 0.21)	0.271

Table 14: Results from models that include an interaction between the intervention indicator and each subgroup variable

		Subgroup × Intervention (differential impact within subgroup level 1 compared to level 0)		Intervention + Subgroup × Intervention (impact within subgroup level 1)	
(1) Subgroup variable; variable levels	**(2) Total n (missing)**	**(3) Effect size (95% CI)**	**(4) p-value**	**(5) Effect size (95% CI)**	**(6) p-value**
FSM status; yes (1), no (0)	1560 (234)	-0.03 (-0.18, 0.11)	0.658	0.11 (0.02, 0.19)	0.015
SEN status; yes (1), no (0)	1560 (234)	-0.07 (-0.22, 0.08)	0.362	0.07 (-0.05, 0.19)	0.232
SENT-R score; ≤40 (1), >40 (0)	1560 (234)	0.05 (-0.17, 0.26)	0.673	0.13 (0.06, 0.21)	<0.001
Cohort; Cohort 1 (1), Cohort 2 (0)	1560 (234)	0.07 (-0.08, 0.21)	0.386	0.14 (0.06, 0.22)	<0.001

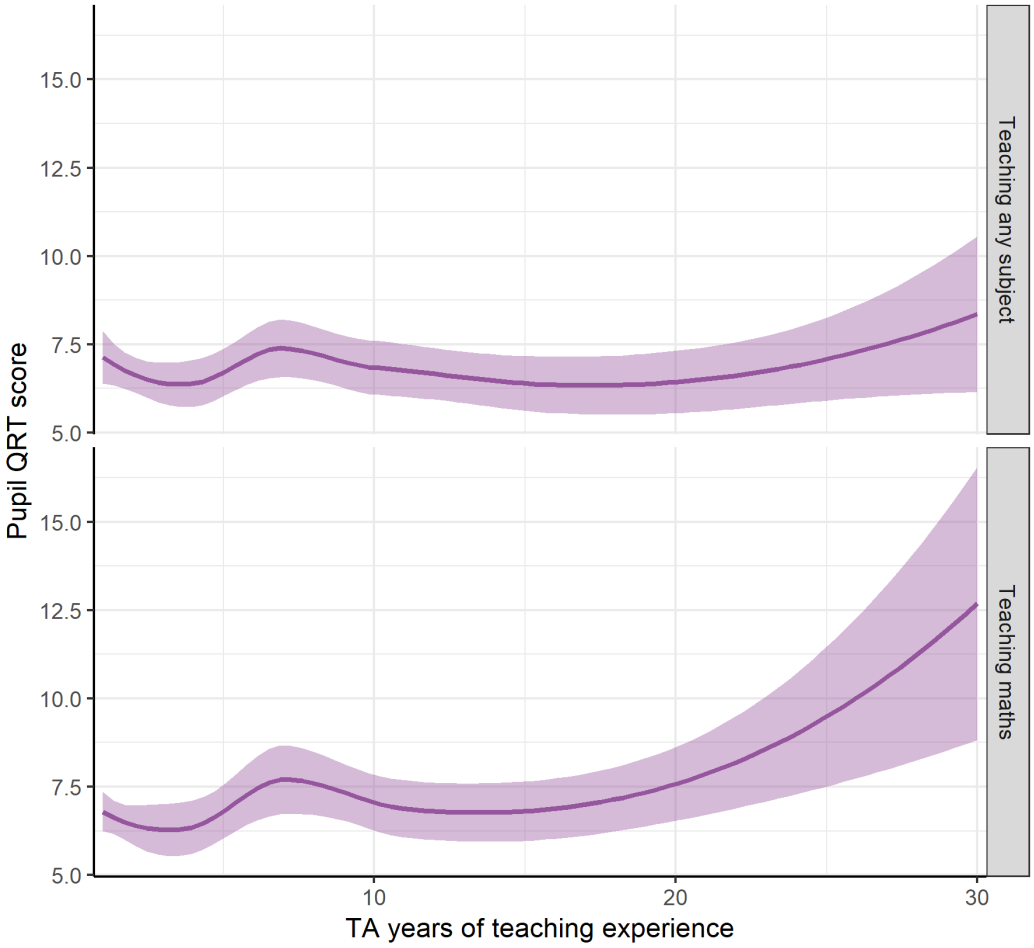
Additional analyses and robustness checks

TA experience and confidence with maths

As we considered it possible that the relationship between endpoint QRT scores and TA years of teaching experience could be non-linear, we first plotted these relationships before proceeding with modelling (Figure 5). These plots did not convincingly demonstrate such a relationship, linear or otherwise, between endpoint pupil QRT scores and years of TA teaching. It could be argued that there is a positive relationship between QRT scores and 20 to 30 years of maths teaching but this trend is based on only a handful of data points in that region of the plot. We therefore proceeded with the default approach of modelling years of TA teaching experience—in maths and in any subject—as linear predictors.

The results from models exploring the moderating effect of TA experience and confidence with maths on pupil QRT scores are shown in Table 15. Each TA-level predictor was included in its own regression model, then all four predictors were included simultaneously in a fifth multivariate model. It can be seen in Table 15 that there was a high proportion of missing data for all models, which reduced the precision of the analysis and increased the potential for bias due to missing data. The estimated impact on QRT scores was close to zero for all predictors and the choice of modelling predictors individually or simultaneously did not make a substantive difference to the results. There is no evidence, based on this analysis, that TA experience and confidence with maths moderates the impact of 1st Class@Number on pupil maths attainment in the QRT.

Figure 5: The relationship between pupil QRT scores and years of TA experience teaching any subject (upper plot) or maths (lower plot)



The relationships are estimated by plotting a LOESS smoothed line, along with a 95% CI (the shaded region), through the raw data points (that is, not dependent on model covariates). Underlying data points are not shown to ensure individual pupils cannot be identified from the figures in this report.

Table 15: Results from five models investigating the impact of TA-level variables on QRT scores: models 1 to 4 include each TA variable as a predictor in its own regression; model 5 includes all four predictors in the same multivariate regression

Model	Total n (missing)	Predictor	Estimate (95% CI)	p-value
1	574 (328)	TA has GCSE maths C+	-0.17 (-1.60, 1.27)	0.821
2	639 (263)	TA years teaching maths	0.03 (-0.04, 0.11)	0.390
3	639 (263)	TA years teaching any subject	0.00 (-0.07, 0.07)	0.994
4	593 (309)	TA maths confidence score *	0.03 (-0.12, 0.18)	0.697
5	542 (360)	TA has GCSE maths C+ TA years teaching maths TA years teaching any subject TA maths confidence score*	-0.39 (-1.92, 1.14) 0.12 (-0.02, 0.26) -0.07 (-0.19, 0.05) 0.04 (-0.11, 0.19)	0.625 0.105 0.285 0.576

* This score had a range of possible values between 6 and 24, with higher values indicating greater confidence.

Sensitivity check—impact of online training

As a sensitivity check, the primary analysis model was restricted to pupils whose TA attended 1st Class@Number 1 training at least partly in person, excluding those whose TA attended training purely online. The effect size in this subgroup was 0.11 (95% CI: 0.04 to 0.18), close to the primary analysis effect size of 0.12. The primary analysis result was therefore mostly unaffected by the inclusion of pupils whose TA attended purely online training, perhaps unsurprising given that there were only 50 such pupils.

Table 16: Impact of the intervention for pupils whose TA attended training in person

	Unadjusted means				Effect size		
	Intervention group		Control group				
Subgroup	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Effect size (95% CI)**	**p-value**
Pupils whose TA attended training in person	739 (99)	6.62 (6.30, 6.94)	799 (93)	6.11 (5.80, 6.41)	1510 (728; 782)	0.11 (0.04, 0.18)	0.002

Sensitivity check—allowing the impact of the intervention to vary between schools

To assess the sensitivity of the primary analysis result to the assumption that the effect of the intervention is fixed across schools, further models were calculated that relaxed this assumption by adding a random slope term to the model.

Likelihood ratio tests were used to test whether the addition of a random slope term to the primary analysis and FSM subgroup models improved model fit (whether $\sigma_E^2 = 0$). These tests produced a p-value of 0.040 ($\chi^2 = 6.42$) in the model for all pupils and a p-value of 0.168 ($\chi^2 = 3.46$) in the model for FSM pupils. This means there is some evidence that the impact of the intervention varies between schools but it is less clear whether this is true for the FSM subgroup. Using the likelihood ratio test to test the significance of a random effect is not an exactly accurate procedure because the null value of zero lies at the boundary of the possible values for a random effect variance. As a result, the p-values reported here will be slightly too high and should be regarded as upper bounds for the correct p-values. The proportion of school-level variance that was due to the differing impact of the intervention between schools, known as Hedges' ω (Hedges and Rhoads, 2010) was 0.22 (see variance parameters in Appendix D).

In the random slope models the effect size among all pupils was 0.13 (95% CI: 0.05 to 0.22) and among FSM-eligible pupils it was 0.12 (95% CI: 0.02 to 0.22), as shown in Table 17. These effects sizes are slightly higher, and the confidence intervals slightly wider, than in the models with only a random intercept. However, the substantive conclusions remain the same: 1st Class@Number 1 was estimated to have a small positive impact on endpoint QRT scores and the data supports a range of small and moderately sized positive impacts to a lesser degree (not including zero impact).

The distribution of school-specific intervention effects is plotted in Figure 6. The effects have been estimated by extracting the 'best linear unbiased predictor' of the random slope for each school and adding this to the overall fixed intervention effect, which was 0.52 on the raw scale (before transformation to the effect size of 0.13 seen in Table 17). According to this plot, a significant minority of schools saw no benefit from the intervention, while others saw an impact several times greater than the overall fixed effect.[27] A range of factors could have caused this wide spread of impact estimates, including differences in pupil-level characteristics between schools and differences in intervention compliance

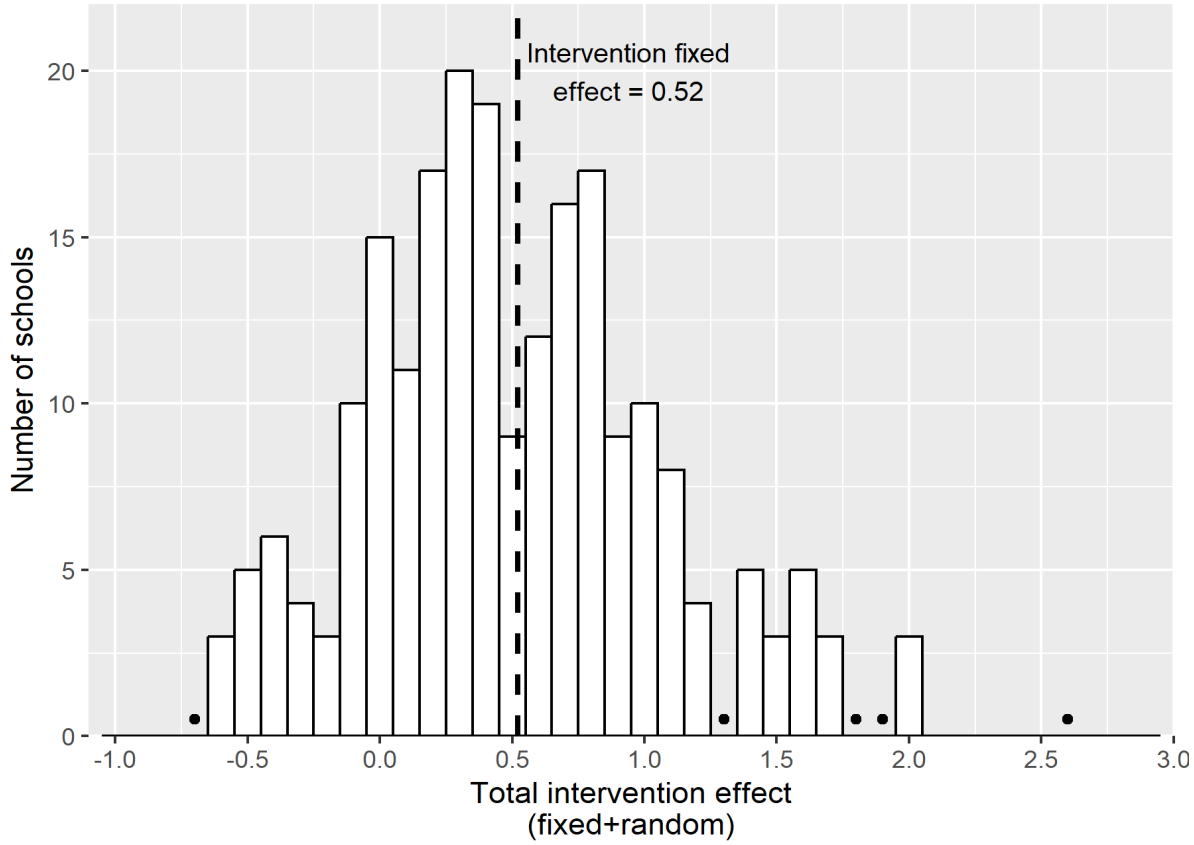
[27] Although these conclusions are suggested by the plot, we advise caution. There is a very large amount of uncertainty around the intervention impact for any one school, each of which provides a sample size of only eight pupils.

or fidelity between schools. In future work it could be informative to investigate why the intervention appears to be more effective in some schools than others, but this is beyond the scope of the current evaluation.

Table 17: Results from sensitivity analyses that add a random slope (school-specific intervention effect) to the primary analysis and FSM subgroup models

	Unadjusted means				Effect size		
	Intervention group		Control group				
Population	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**Total n (intervention; control)**	**Effect size (95% CI)**	**p-value**
All pupils	789 (113)	6.64 (6.33, 6.95)	799 (107)	6.11 (5.80, 6.41)	1560 (778; 782)	0.13 (0.05, 0.22)	0.003
FSM pupils	500 (74)	6.42 (6.04, 6.8)	512 (67)	5.88 (5.5, 6.26)	993 (494; 499)	0.12 (0.02, 0.22)	0.021

Figure 6: Histogram showing the distribution of school-specific intervention effects: the total intervention effect for each school is estimated by summing the overall fixed effect and the 'best linear unbiased predictor' of the school's random effect



Each histogram bar counts the schools with an intervention effect +/- 0.05 of the number given on the x axis. For example, the bar above '1.0' includes schools with an intervention effect between 0.95 and 1.05.
Black dots on the histogram represent school counts greater than zero but less than three. These have been supressed to ensure individual schools cannot be identified from the figures in this report.

Analysis in the presence of non-compliance

The impact of compliance with the intervention on QRT scores was investigated using four different definitions of 'compliance', summarised below.

(1) Pupil attended all 30 sessions.

(2) Pupil attended at least five of the six sessions in each of the five topic areas, implying at least 25 of the 30 sessions attended in total.

(3) Pupil attended at least five of the six sessions in each of the five topic areas, as for (2). Additionally, the pupil's TA attended all four of the initial training sessions.

(4) Pupil attended at least four of the six sessions in at least four of the five topic areas, implying at least 16/30 sessions attended in total.

Compliance definition (4) was added after the SAP was published to explore the sensitivity of results for definitions (1) to (3) to the exclusion restriction (further described below). As shown in Table 18, only 40% of intervention pupils met compliance definition (1), more than half met definitions (2) and (3) (61% and 54%) and 95% met definition (4). These figures are only for pupils included in CACE analysis so do not exactly match those reported in the IPE. Intervention pupils with missing compliance information (some schools did not provide session attendance data) were not included in modelling, meaning that the number of pupils included in the compliance analysis (N = 1,510 or 1,509, depending on the model)[28] was lower than the primary analysis (N = 1,560).

A small number[29] of control pupils received one or more 1st Class@Number 1 sessions due to either to an administrative error by schools or their choice to switch pupils (see Usual Practice section). If this resulted in them meeting a compliance definition, they were analysed accordingly: their outcome in the first stage of the two-stage least squares regressions was entered as '1', not '0'. The impact of these few pupils on the compliance results is likely to be minimal.

The Complier Average Causal Effect (CACE) estimates obtained from instrumental variable models (one for each compliance definition) are given in Table 18 where they have been converted into effect sizes. CACE effect sizes for compliance definitions (1) to (3) were substantially larger than that obtained using an intention-to-treat approach in the primary analysis, particularly for compliance definition (1). The CACE effect sizes in the subgroup of FSM pupils were similar to the corresponding effect sizes among all pupils. These results only apply amongst 'compliers' (pupils that met the compliance definition)[30] and may not extrapolate to all pupils in the sample.

Although the large CACE effect sizes for compliance definitions (1) to (3) seem promising for the intervention, these results should be interpreted with extreme caution. This is because instrumental variable analysis relies on the exclusion restriction assumption, which stipulates that there can be no benefit of 'partial treatment' (that is, receiving some but not all of the intervention) below the selected compliance threshold. For instance, for compliance definition (1) it is assumed intervention pupils receiving fewer than 30 sessions gain no benefit from 1st Class@Number 1, even if they attended most of the sessions. We consider it unlikely that this assumption fully holds for compliance definitions (1) to (3), which could result in upwardly biased CACE effect sizes.

The CACE effect size for compliance definition (4) is likely to be approximately correct in terms of the exclusion restriction (as explained in the Methods section). Therefore, it serves as a lower bound for the unbiased CACE effect sizes for compliance definitions (1) to (3). We are reasonably confident that the unbiased CACE effect size for each of these

[28] For control pupils, we collected less detailed compliance data: only total number of sessions attended, without information about which topics they were in. Control pupils could in principle meet the compliance definitions (as discussed in the following paragraph) but for one control pupil with partial compliance it was not possible to establish whether they met compliance definitions (2) and (3), thus the discrepancy of one pupil between model Ns.

[29] Less than ten but greater than zero pupils. The exact number is supressed to prevent individual pupils being identified from these results.

[30] This is a slight simplification of the exact definition of 'compliers', which would need to account for 'always-takers', given that noncompliance is technically two-sided in this case. We consider the simplified version adequate for understanding the results presented here.

three definitions is greater than 0.15. However, there remains a wide margin of uncertainty around the true CACE effect sizes for compliance definitions (1) to (3), especially after also accounting for sampling uncertainty.

Results from the stage one regression of each instrumental variable analysis are shown in Table 19. The primary purpose of these results is to ascertain whether the instrumental variable (in this case, random allocation) is sufficiently correlated with the compliance indicator to allow for valid inference from the model. There is no cause for concern as the F-statistics range from 308 to 8,489, which is well above any minimum value suggested for valid inference in the instrumental variable literature (Lee et al., 2022).

Table 18: CACE estimates from stage two of the instrumental variable two-stage least squares regressions

Group	Compliance definition	Total n (missing)	N (%) of intervention pupils meeting compliance definition	CACE effect size (95% CI)	p-value
All pupils	(1) Pupil attended 30/30 sessions	1510 (284)	292 (40)	0.36 (0.13, 0.59)	0.002
	(2) Pupil attended 5 of the 6 sessions in all topics	1509 (285)	445 (61)	0.24 (0.09, 0.39)	0.002
	(3) Pupil attended 5 of the 6 sessions in all topics and TA attended first 4 training sessions	1509 (285)	391 (54)	0.27 (0.10, 0.44)	0.002
	(4) Pupil attended 4 of the 6 sessions in 4 of the 5 topics	1510 (284)	689 (95)	0.15 (0.06, 0.25)	0.002
FSM pupils	(1) Pupil attended 30/30 sessions	959 (191)	181 (39)	0.33 (0.03, 0.63)	0.030
	(2) Pupil attended 5 of the 6 sessions in all topics	958 (192)	273 (59)	0.22 (0.02, 0.42)	0.028
	(3) Pupil attended 5 of the 6 sessions in all topics and TA attended first 4 training sessions	958 (192)	237 (52)	0.25 (0.03, 0.48)	0.028
	(4) Pupil attended 4 of the 6 sessions in 4 of the 5 topics	959 (191)	439 (95)	0.14 (0.01, 0.26)	0.030

Table 19: Results from stage one of the instrumental variable two-stage least squares regressions

Group	Compliance definition	F test statistic (p-value)	Intervention indicator estimate (95% CI)	p-value
All pupils	(1) Pupil attended 30/30 sessions	514 (<0.001)	0.37 (0.34, 0.41)	<0.001
	(2) Pupil attended 5 of the 6 sessions in all topics	1148 (<0.001)	0.57 (0.54, 0.60)	<0.001
	(3) Pupil attended 5 of the 6 sessions in all topics and TA attended first 4 training sessions	892 (<0.001)	0.51 (0.47, 0.54)	<0.001
	(4) Pupil attended 4 of the 6 sessions in 4 of the 5 topics	8489 (<0.001)	0.91 (0.89, 0.93)	<0.001
FSM pupils	(1) Pupil attended 30/30 sessions	308 (<0.001)	0.36 (0.32, 0.40)	<0.001
	(2) Pupil attended 5 of the 6 sessions in all topics	665 (<0.001)	0.55 (0.51, 0.59)	<0.001
	(3) Pupil attended 5 of the 6 sessions in all topics and TA attended first 4 training sessions	517 (<0.001)	0.48 (0.44, 0.52)	<0.001
	(4) Pupil attended 4 of the 6 sessions in 4 of the 5 topics	5674 (<0.001)	0.91 (0.89, 0.94)	<0.001

Missing data analysis

N = 237 randomised pupils (13.2%) were not included in the primary analysis due to missing data from all sources (intervention N = 126, 13.9%; control N = 111, 12.4%). This was mainly due to missing outcome data: N = 209 pupils (11.6%) had a missing endpoint QRT score (intervention N = 115, 12.7%; control N = 94, 10.5%). There is no known reason to think the slight imbalance in attrition rates between the control and intervention is indicative of a systematic difference: figures in Table 20 show the imbalance is largely due to pupils leaving the school, which is unlikely to be connected to the intervention. N = 40 pupils (2.2%) had a missing QRT baseline score (intervention N = 20, 2.2%; control N = 20, 2.2%) and N = 10 pupils (0.6%) had missing FSM status on the NPD (intervention N = 7, 0.8%; control N = 3, 0.3%).

Multilevel logistic regressions were run in which the outcome was whether a pupil's baseline QRT score was missing (1) or not missing (0) (Appendix G). Similarly, regressions were run in which the outcome was missingness of endpoint QRT scores (Appendix G). These regressions were performed for both all pupils and FSM pupils, for a total of four models. All p-values were above 0.05, so no predictors were shown to be associated with missingness from either the primary analysis variables (baseline and endpoint QRT scores, randomisation group, FSM) or the auxiliary variables (SEN, SENT-R score, establishment type, school FSM quintile).

There were some technical issues with estimating the multilevel logistic regression models: all four models produced convergence warnings, indicating that the maximum likelihood solutions may not have been reached. These warnings were probably caused by 'quasi-complete separation': due to the rarity of the missing QRT score outcome, some cells in the cross tables created by the outcome and certain categorical predictors were empty. Following further

investigations,[31] we decided that the model for missing endpoint QRT scores (among all pupils) was fairly reliable but the other three models were not. We could not conclude that any observed variables predicted missingness of QRT scores, either because p-values were above 0.05 or because there was insufficient information to draw a conclusion (due to the issues described). This does not prove that a complete case analysis is unbiased, only that if it is biased, there is no evidence that conditioning on the available variables would remove the bias.

Given that observed variables did not explain the missingness in the QRT outcome, we conducted further sensitivity analysis allowing the possibility that outcome data was missing not at random. The complete approach is explained in the Methods section but, to recap, three missing data patterns were considered possible:

- 'MAR'—missing at random conditional on primary analysis covariates;
- 'MNAR balanced'—outcomes are missing not at random: all pupils with a missing outcome have their outcome imputed as one standard deviation lower than those with an observed outcome; this is 'balanced' in the sense that both intervention and control pupils with missing outcomes are affected equally; or
- 'MNAR unbalanced'—outcomes are missing not at random: intervention pupils with a missing outcome have their outcome imputed as if they were control pupils (that is, as if they received no benefit from the intervention); this is 'unbalanced' in the sense that only intervention pupils are affected.

The missing data pattern assumed for a pupil with a missing outcome depended on which of the six absence codes they had on their record for the endpoint QRT (see the first column of Table 20). Under three sets of assumptions (scenarios one to three in Table 20) multiple imputation was performed for pupils with missing data to investigate whether the primary analysis result was robust to these missing data scenarios.

A summary of each missing data scenario modelled and the intervention impact estimate for that scenario is given in Table 20. In the most conservative scenario that we considered possible (scenario one) the estimate derived from the multiple imputation procedure was 0.40 (95% CI: 0.09 to 0.71). This was a 23% reduction on the primary analysis estimate of 0.52 (95% CI: 0.22 to 0.82) and, like that estimate, was incompatible with zero intervention impact. Given that this was a 'worst case' scenario, we consider the primary analysis result to be fairly robust to any bias that could have been introduced by missing outcome data. Note that these estimates are on the raw QRT scale and have not been converted into effect sizes.

The choice to assume that the N = 95 pupils who were absent for the endpoint QRT had MAR outcomes for all three scenarios might be questioned, as persistent absence could be correlated with lower attainment. However, the number of sessions attended by intervention pupils that were absent for the endpoint QRT (mean 25.4) was only slightly lower than for intervention pupils that had a recorded QRT outcome (mean 26.5). This suggests their absence on the day of testing was not strongly correlated with either (1) how much of the intervention they received or (2) their overall absence rate.[32] It is also worth noting that the N = 39 intervention pupils that left the school during the trial period averaged only 8.1 sessions, making the 'MNAR unbalanced' assumption—that they received little or no benefit from the intervention—likely to be close to the truth.

[31] Models were refit using the default range of optimisers available in the `lme4` package, see https://search.r-project.org/CRAN/refmans/lme4/html/allFit.html. For the missing endpoint QRT score model (all pupils), results were nearly identical between optimisers, and some of the optimisers converged (no warning message). For the other three models, results did not agree between the optimisers and all optimisers produced convergence warnings. Extremely high standard errors were observed for the school type and FSM quintile predictors in particular (where these clearly inaccurate numbers are replaced with 'not estimable' in the appendix tables).

[32] This second point is more speculative and would be better answered by looking directly at pupil's overall absence rates, but that information was not available.

Table 20: Sensitivity analyses in which missing QRT outcomes are imputed under different assumptions; the assumptions made about the values of missing outcomes depend on the reason why that outcome is missing

Reason for missing QRT outcome	Total n (intervention; control) *	Scenario 1	Scenario 2	Scenario 3
School withdrawal	16 (8; 8)	MNAR unbalanced	MNAR unbalanced	MAR
Pupil left school	69 (39; 30)	MNAR unbalanced	MNAR unbalanced	MAR
Pupil did not want to take test	<10 (<10; <10)	MNAR balanced	MAR	MNAR balanced
Pupil present but excluded	11 (<10; <10)	MNAR balanced	MAR	MNAR balanced
Pupil absent on day of test	95 (49; 46)	MAR	MAR	MAR
Pupil withdrawal	<10 (<10; <10)	MAR	MAR	MAR
Intervention estimate (95% CI)		0.40 (0.09, 0.71)	0.43 (0.12, 0.73)	0.51 (0.23, 0.8)

* Pupil counts below ten but above zero are supressed to prevent individual pupils being identified from these figures. The exception is for the school withdrawal numbers: eight pupils per school is known anyway due to the trial design.

Estimation of the intracluster correlation coefficient (ICC)

The ICCs from a model with no covariates (the unconditional ICC) and from the primary analysis model (the conditional ICC) are given below in Table 21.

Table 21: The unconditional and conditional intracluster correlation coefficients

Unconditional ICC			Conditional ICC		
Within-school variance σ_W^2	Between-school variance σ_B^2	ICC	Within-school variance σ_W^2	Between-school variance σ_B^2	ICC
12.52	6.86	0.35	9.05	6.51	0.42

Implementation and process evaluation results

The following sections describe the IPE results, beginning with who took part in the intervention and what they did prior to the intervention (BAU) before going on to describe views about and experiences of the intervention. Subsequent sections then consider the perceived outcomes for different participant groups before finishing with views about scale-up and sustainability. Many of our research questions cross-cut multiple domains; we have included labels indicating the research questions covered in the domain headings. We note that our results may be subject to response bias as schools that liked the intervention are more likely to complete the surveys and/or participate in case studies. However, as our survey response rates were high, we judge this risk to be low.

Reach (RQ1.1, 1.2)

IPE RQ1: Who was selected to participate in the intervention and how were they chosen?

To participate in 1st Class@Number 1, schools were required to select a Link Teacher to oversee and support the implementation of the intervention in the school and a TA to receive the training and deliver it. Through the surveys and case study interviews, the IPE sought to understand the considerations when schools selected these members of staff and the processes schools followed, guided by selection criteria, for nominating pupils for the trial.

Link teacher selection

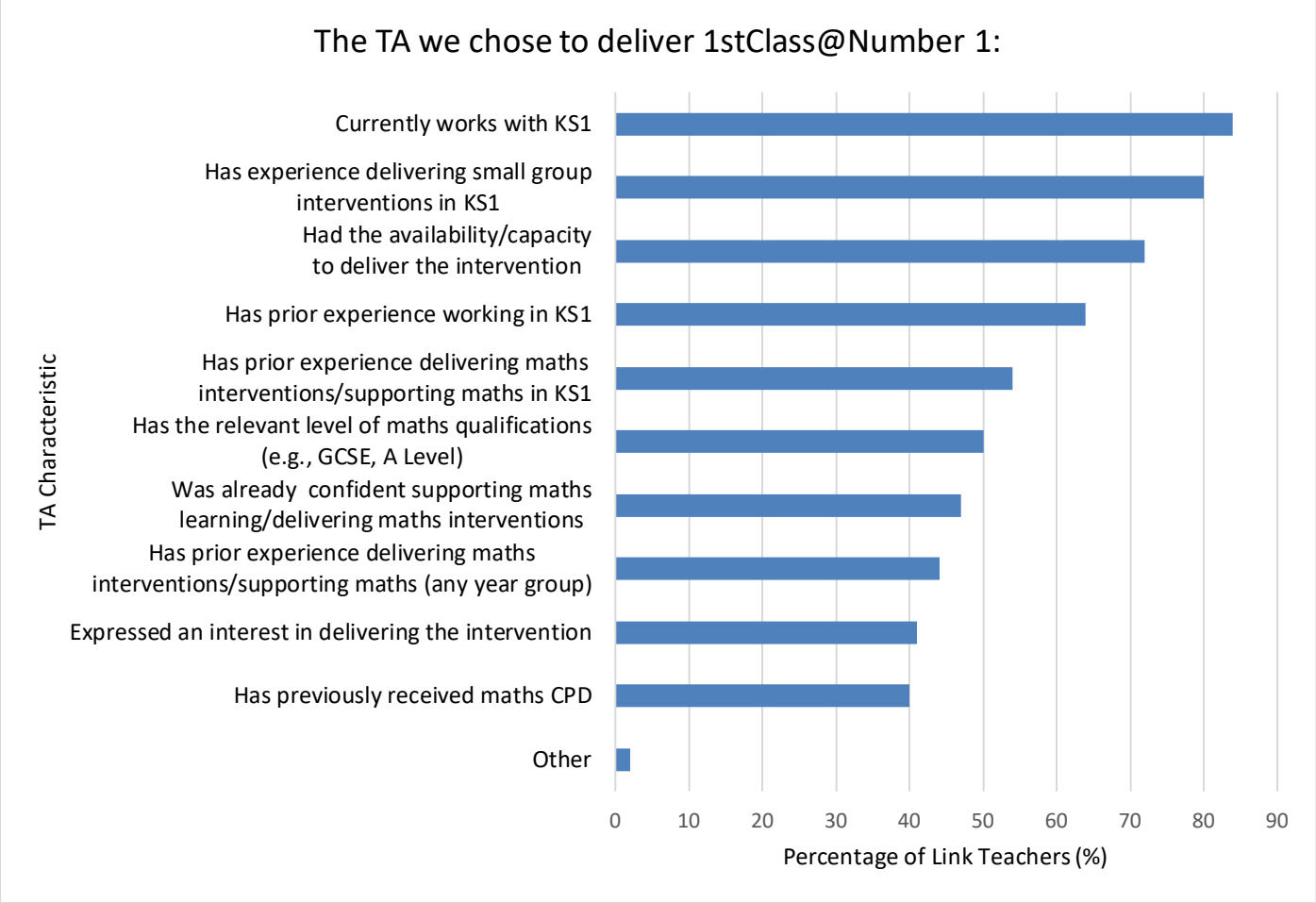
Selection of the Link Teacher is an important consideration when schools sign up to the intervention as this member of staff is expected to provide strategic leadership of 1st Class@Number 1 in the school so ideally holds a role which facilitates this. The delivery team also advises that Link Teachers are experienced at providing maths teaching support as throughout delivery of the intervention they are expected to support the TA with implementation, such as through supporting the planning and reviewing of sessions. Link Teachers must also commit to attending the first and last half-day training sessions alongside the TA. The survey asked Link Teachers to comment on their role in school. The question allowed multiple responses and almost half (49%) of the 174 respondents held multiple roles:[33] Most tended to be the school maths lead (52% of survey respondents) and/or were a class teacher (47%). Around a third were a member of the senior leadership team (SLT): 3% reported being the headteacher; 28% a deputy or assistant headteacher. Those reporting a head of key stage or head of year responsibility (24%) were mostly working within the Early Years Foundation Stage (EYFS) and/or Key Stage 1. A smaller proportion indicated that they were the SENCo (11%) or held another role (5%) such as curriculum, teaching, and learning lead or intervention teacher. Link Teachers in case study schools reported that the reason they were selected for their role by the headteacher or put themselves forwards to take on the role was due to their responsibilities as maths lead, intervention lead, Key Stage 1 phase lead or because they were part of the school's SLT (as reported in the survey). There were cases where, in addition to these responsibilities, the Link Teacher also taught Year 2 pupils.

Teaching assistant selection

Through a multiple-response survey question, Link Teachers were asked about the TA they had selected to deliver 1st Class@Number 1. The majority of the 174 survey respondents reported choosing someone who had experience relevant to the age-group of the intervention, including currently working in Key Stage 1 (84%), and experience delivering small-group interventions in Key Stage 1 (80%) (see Figure 7). Availability and capacity to deliver the intervention also appeared to be important: nearly three-quarters of Link Teachers (72%) reported this as a key issue. The programme handbook for 1st Class@Number 1 states that the TA should have 'successful experience supporting pupils' mathematics in the appropriate age phase'; however, only around half of Link Teachers (54%) reported that the TA they chose to deliver the intervention had prior experience delivering maths interventions with Key Stage 1. This suggests that for around half of schools whose Link Teacher responded to the survey, the TA delivering the intervention may not have had prior experience supporting pupils in maths in the relevant year group. That said, the majority of Link Teachers (95%) felt they selected the most appropriate TA for delivery.

[33] Fifty-five (31%) Link Teachers held two roles in school; 24 (14%) held three roles and seven (4%) four roles.

Figure 7 7: TA selection to the trial



Data from Link Teacher survey: 'the TA we chose to deliver 1stClass@Number 1' (N = 174). Participants could select more than one response. Items from the question have been reordered to be presented in descending order.

Qualitative findings from case study interviews with Link Teachers broadly align with survey findings. Link Teachers reported that they had selected the TA because of their experience, both in delivering interventions and working within Key Stage 1. This meant they had experience working with Key Stage 1 pupils in a small group setting. Across most case study schools, the TA who delivered the intervention was a Year 2 TA, which meant they already had a relationship with the intervention pupils. Link Teachers saw this (that is, working with a known member of staff) to be an important factor for supporting pupil engagement with the intervention. Link Teachers reported confidence in the TA to implement the intervention to a high standard with fidelity and felt their level of experience and skills would support them to make ability-based adaptations where required to support pupils.

The TA survey sought to further understand their experience, both in their role and of delivering interventions (Table 22). On average (median), TAs who delivered the intervention had ten years' experience—inter-quartile range (IQR): 13; quartile 1 (Q1): 5.0; quartile 3 (Q3), 18.0—as a TA. All of the 187 respondents indicated that they had experience delivering interventions in any subject (that is, not specifically maths interventions) to Key Stage 1 pupils. This, however, contrasts with the 80% of Link Teachers who indicated that the TA chosen had such experience. On average (median), TAs reported that they had four years' experience (IQR: 8; Q1: 2.0; Q3: 10.0) delivering interventions in any subject to Key Stage 1 pupils as part of their TA role, however, a fifth (19%) indicated that they had less than one year experience, which may account for the discrepancy between reports across Link Teachers and TAs. Similarly, all TAs reported that they had experience delivering maths interventions to Key Stage 1 pupils but, again, this contrasts with Link Teacher reports, 54% of whom indicated their selected TA had such experience. On average, TAs had a median of three years' experience (IQR: 7; Q1: 1.0; Q3: 8.0) specifically delivering KS1 maths interventions yet just over a third (37%) reported less than one years' experience. If Link Teachers are not considering TAs with less than a years' experience as having 'prior experience' when answering the survey question, this may go some way to explaining the differing percentages across the Link Teacher and TA surveys. Most TAs who completed the survey (79%) reported that they had achieved a

Grade C/4 or above in their maths GCSE or equivalent, with the remaining proportion reporting they had not achieved this grade (12%) or were unsure (10%).

Table 22: TA responses to knowledge and experience survey questions

Question	Median	Inter-quartile range [Q1; Q3]
How long have you been a teaching assistant?	10 years	13 years [5; 18]
How many years' experience do you have delivering maths interventions to Key Stage 1 (KS1) pupils?	3 years	7 years [1; 8]
How many years' experience do you have delivering interventions to KS1 pupils in any subject?	4 years	8 years [2; 10]

The survey asked TAs to comment on their confidence in their maths abilities overall, and across specific skills. The majority reported feeling confident across items related to mental maths abilities, solving maths problems, using maths in daily life, and in their belief that they would be able to understand the content delivered in maths CPD (92% strongly agreed or agreed they were confident with each individual item). A slightly smaller proportion believed they were good at maths (87%) but overall, 90% reported they were confident in their own maths abilities.

TAs interviewed in case study schools reported that the schools' Link Teacher, or a member of SLT, had approached them to deliver 1st Class@Number 1 and that they had been happy to receive the training. Some also commented that they had previously expressed an interest in receiving CPD to develop their pedagogical knowledge, so this had been a welcome opportunity.

Selection of pupils

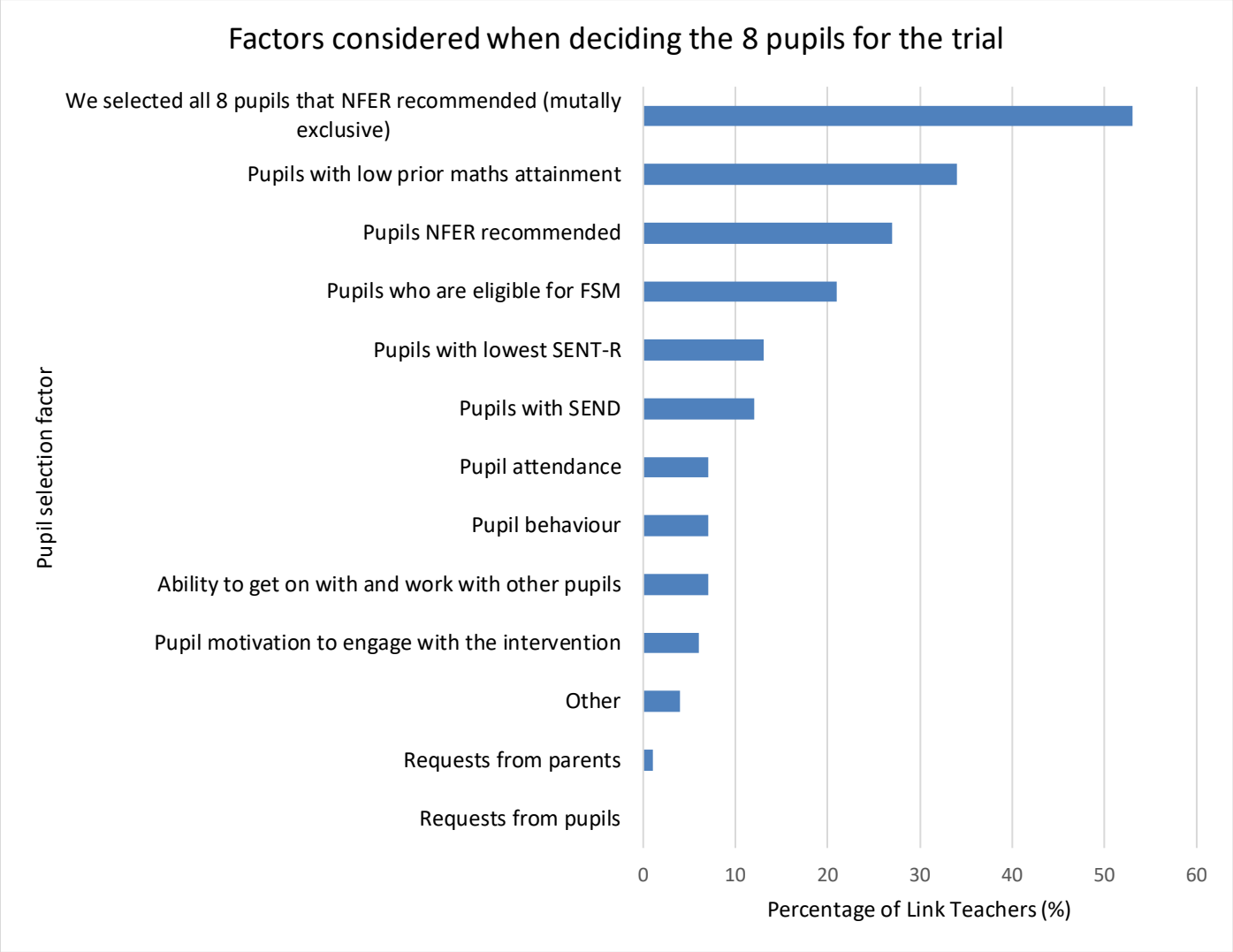
In this trial, schools nominated ten to 16 pupils to be screened, using the SENT-R, for trial eligibility. Based on the results of this screening, NFER then recommended eight pupils per school to take part, prioritising FSM pupils with eligible scores ahead of non-FSM pupils with eligible scores. Ahead of randomisation, schools had the opportunity to switch out any non-FSM pupils they did not feel were well suited for the trial. To maintain the trial's focus on FSM pupils, schools were asked to not switch out FSM pupils recommended by NFER. However, since this was an effectiveness trial, schools could make this switch if they strongly felt that non-FSM pupils were more likely to benefit than the FSM pupils recommended by NFER. Schools were provided with guidance on how to select pupils for screening and around switching eligible pupils recommended for the trial for other eligible pupils who had been screened. Of the 1,797 NFER-recommended pupils (1,150 FSM pupils and 637 non-FSM pupils, 10 pupils had FSM status missing), 48 schools switched a total of 72 pupils (4%): ten FSM pupils recommended by NFER were switched by schools for screened non-FSM pupils and eight were switched for other screened FSM pupils; the remaining 56 were non-FSM pupils switched for other non-FSM pupils.

In case study schools, the main reasons reported by Link Teachers for selecting pupils for initial screening were pupils not meeting age-related expectations (ARE) at the end of Year 1, lacking confidence in maths, and being eligible for Pupil Premium funding or free school meals. For the most part, case study schools had stuck with the eight pupils recommended for randomisation by NFER. Exceptions included switching out pupils who had severe special educational needs or disabilities (SEND) or were considered to no longer need the intervention. Link Teachers interviewed in case-study schools reported that they were satisfied with the guidance they had been provided for this screening and selection process.

The findings from the case study are consistent with those from the Link Teacher survey. This survey asked Link Teachers to identify the factors they had considered when deciding on the final eight pupils who would participate in the trial. Just over half of Link Teachers who completed the surveys (N = 93, 53%) reported selecting all eight pupils recommended by NFER (note that this option was mutually exclusive). However, this proportion of respondents appears low given that only 4% (N = 72, across 48 schools) of recommended pupils were switched by schools. It is possible that

schools may have considered other factors but then chose the NFER recommended selection, which may have influenced how they responded to the question. Among Link Teachers who reported that they had made changes to NFER's recommended selection, the main considerations when selecting pupils for the intervention were: pupils with low prior maths attainment (34%); pupils NFER recommended (27%) and pupils eligible for FSM (21%). Figure 8 shows the percentage of Link Teachers who selected each response regarding their selection of the eight pupils for the trial.

Figure 8: Factors considered by schools when selecting pupils for the trial



Data from Link Teacher survey: 'What factors did you consider when deciding on the final eight pupils who would participate in the trial?' (N = 173). Items from the survey question have been reordered to be presented in descending order.

Across the majority of items, the analysis found little to no differences in the reporting of these factors across Cohorts 1 and 2. The only notable differences were in relation to pupil behaviour (a consideration of 12% of Cohort 2 Link Teachers compared to 6% of Cohort 1 Link Teachers) and pupils' ability to get on with and work with other pupils (a consideration of 13% of Cohort 2 Link Teachers compared to 5% of Cohort 1 Link Teachers).

Usual practice (RQ5.1, 5.2, 5.3)

RQ5: What was business as usual (BaU) and did this change during the intervention?

To understand usual practice in trial schools, at the end of the trial, schools completed an intervention record for each of the eight pupils involved, detailing the maths interventions they had received over the trial period. The research team also explored schools' usual approach for supporting Year 2 pupils in maths through the case study interviews with Link Teachers.

Usual practice in intervention schools

Link Teachers in case study schools reported that a range of whole-class teaching approaches combined with other small-group interventions formed their usual practice for teaching maths in Year 2. Link Teachers reported that their class teaching followed schemes or programmes aligned with a mastery model of teaching, such as from the National Centre of Excellence in the Teaching of Mathematics (NCETM). Larger primary schools were more likely to use ability streaming in maths. Case study data and maths interventions records showed that interventions used with pupils identified as needing additional support in maths included programmes such as Numbers Counts, Ark Maths, Plus One, and Numicon, as well as pre-teaching or same-day intervention based on pupils' current performance in maths.

Reasons for engaging in 1st class@Number 1

Link Teachers interviewed reported a range of reasons for choosing to engage with the 1stClass@Number 1 trial, which were dependent on whether they had previously engaged with or had experience of the intervention. Those who had engaged with the intervention before[34] reported that they had enjoyed delivering the programme and had been pleased with its impact on pupils' progress. Link Teachers new to the intervention commented that across the school, their teaching practice is evidenced-informed and spoke of the programme's reputation and the research evidence for improving results and confidence in maths which they hoped to emulate. This aligned with their schools' maths priorities for boosting attainment and increasing the percentage of pupils achieving ARE.

Among schools not already using formal maths interventions in Year 2, 1stClass@Number 1 was seen to provide a structure and focused topic areas for supporting lower-ability pupils to secure the foundations of maths—elements Link Teachers perceived to be sometimes missing from pre-teaching or same-day interventions that were responsive to pupils' current needs. Link Teachers reported that 1stClass@Number 1 clearly aligned with their Year 2 maths curriculum and the NCETM's 'ready to progress' criteria. Link Teachers also commented that the subsidised cost of the intervention through the trial (£200 rather than the usual £1,100) had made the intervention affordable—something it had not been in previous academic years. This aligns with findings reported in the Costs and Sustainability section of this report, in which 76% of Link Teachers reported that they would have been unable or unwilling to pay £1,100 for the intervention,

Additional maths support for intervention and control pupils during the trial[35]

Intervention records indicated that a small number (fewer than ten of 778)[36] of control pupils we received maths intervention records for received the 1stClass@Number 1 intervention (these pupils received between 23 and 30 sessions). These pupils were spread across three different schools and received the intervention either in error or due to the school's decision to switch intervention and control group pupils due to 'unforeseen circumstances'. Note that switching was not discussed or agreed with the research team. No other control pupils for whom we received completed maths intervention records received any 1stClass@Number 1 sessions. Therefore, it appears that contamination of the control group in terms of receiving the intervention programme was very limited.

Among pupils for whom we received maths intervention records (N = 1,561), 15% (N = 228) were missing data about whether they received any additional forms of support during the trial.[37] Of the remainder (N = 1,333), 226 were reported to have received additional maths interventions or support during the trial (i.e. in addition to 1stClass@Number 1 for intervention pupils); this was : 15% of intervention pupils (102/675) and 19% of control pupils (124/658). Statistical significance indicated that a slightly higher proportion of control group pupils received additional maths/interventions support than the intervention group (p =0.042);[38] This maths support included formal interventions such as Number

[34] To be eligible to participate in the trial, schools must not have had a TA currently working in the school who had already completed 1stClass@Number training, and the school must not have delivered the intervention since 1 September 2019.

[35] Note that all pupil-level data reported in the IPE is for pupils for whom we received implementation data (for example, registers, maths intervention records). This differs from the impact evaluation, which is based on pupils who had baseline and endpoint QRT data.

[36] 893 control pupils participated in the trial, so we received maths intervention records for 87% of control pupils.

[37] For intervention pupils, such support was in addition to 1stClass@Number 1.

[38] Specifically, a multilevel logistic regression was used where the outcome was whether a pupil received extra support ('1') or did not ('0') and the only predictor was randomisation group. The odds ratio was 0.53 (95% CI: 0.29 to 0.97).

Sense Maths, Addition Facts Fluency, and the Mastering Number Programmes, as well as other types of small-group support on specific topics or as pre-teaching or additional maths support.

In total, 226 pupils across the intervention and control groups were reported to have received such additional maths support. Among these pupils, we received data about who delivered the main type of additional maths support for 192 pupils (85%) across the intervention (N = 82) and control (N = 110) groups. The data showed that in the vast majority of cases support was delivered by a TA (control N = 80; intervention, N = 53) or the pupils' class teacher (control, N = 27; intervention, N = 26). In the remaining three cases in each group, support was provided by another teacher at the school.

Pupils in both groups received this support between one and five days per week and sessions lasted from less than ten minutes to more than 40 minutes. The median session duration in both the intervention and control groups was 20 minutes (IRQ = 20 minutes [ten minutes, 30 minutes] and the median number of sessions per week was three (IRQ = 3 [1, 4]). However, in both groups it was most commonly reported that pupils received one additional maths session per week (control, N = 32/113; intervention, N = 28/95) and that additional maths sessions lasted for up to 20 minutes (control, N = 35/110; intervention, N = 27/91), suggesting that in the small number of cases where pupils did receive additional maths support, it was similar in the intervention and control groups.

Information about the number of weeks of additional maths support pupils received was recorded for a total of 190 pupils (84% of the pupils recorded as having received additional support) again with more data recorded for control group pupils (85%, N = 105) than intervention group pupils (83%, N = 85). This data indicated that 45 control pupils and 40 intervention pupils received additional maths support lasting more than ten weeks, with around half of these receiving support for 20 weeks or more (control, N = 23/40; intervention, N = 25/40). This means that overall, among pupils for whom data about additional maths interventions was recorded, 7% of control pupils (45/658) and 6% of intervention pupils (40/675) received additional maths support for longer than the ten-week duration of 1stClass@Number 1.

Link Teachers in each of the case study schools reported that the four pupils randomised to the intervention group in their school had not received any other interventions alongside 1stClass@Number 1. They reported being conscious of the trial and wanting to ensure the impact of the programme upon pupil progress was clear. However, as reported above, there were some instances of case study schools delivering their business-as-usual maths interventions to both intervention and control group pupils during the trial.

RQ2: To what extent was fidelity to (a) the intended training design, and (b) the intended intervention delivery achieved?

Compliance and Dosage (RQ2.1, 2.2, 2.3)

TA attendance at training sessions

The ideal training model for 1stClass@Number 1 is that the same TA from each school attends all six training sessions face to face or, in the case of the online group, online. The first four training days (five sessions) focus on the intervention content, meaning they are the training sessions that influence intervention delivery. The fifth day (sixth session) is a reflection session after delivery is complete. To support schools where the TA missed a training session, online catch-up sessions were provided by EHU the week after the original training session was delivered. This meant that TAs could have attended the first five sessions via a mixture of face to face and online catch-up sessions and still received training in the full intervention content. Attendance by the same TA at the first five training sessions (either in-person or as a catch-up) was therefore the key measure the research team used to understand training compliance for this trial.

Among the 226 schools that participated in the trial, 83% (N = 188) had the same TA attend the first four training days (five sessions) in person or as an online catch-up. However, only 70% (N = 159) of TAs attended the first five training sessions as planned (that is, with no catch-up sessions). This suggests that the provision of online catch-up sessions was a valuable addition to the programme as it allowed an additional 29 TAs (13%) to complete the first five training sessions. The vast majority of schools used the same TA to deliver the intervention in school throughout the delivery period, however, eight schools needed to change the TA delivering the intervention part-way through delivery due to

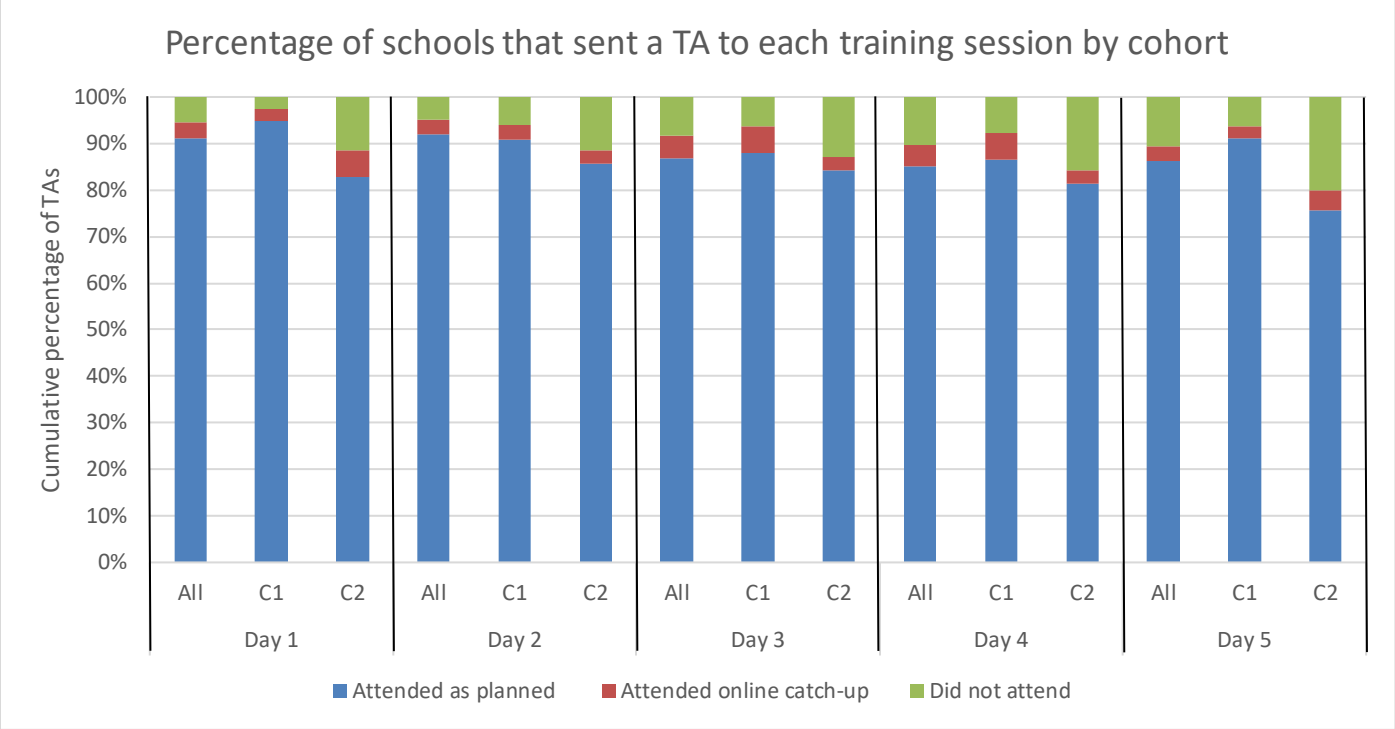
long-term staff sickness or staff turnover. This meant that for these eight schools, training attendance was split between two TAs[39].

Fewer schools (79%, N = 178) were able to send the same TA to all six training sessions either in person or online and only 63% (N = 142) of TAs were able to attend all six training sessions as originally planned. Training attendance at all six sessions as planned was higher in Cohort 1 (66%, N = 103) compared with Cohort 2 (56%, N = 39). Similarly, training attendance at the first five sessions as planned was higher in Cohort 1 (71%, N = 111) compared with Cohort 2 (69%, N = 48). Indeed, attendance at each individual session was also higher in Cohort 1 than Cohort 2, although there was a tendency for attendance to decline towards the end of the programme in both cohorts (see Figure 9). Reasons for non-attendance at in-person training included TA illness, mandatory training events (for example, safeguarding training) taking place in school that clashed with the training, and TAs covering teacher absence.

Link teacher attendance to relevant training sessions

Link Teachers were only required to attend the first and last training sessions (first and sixth). Across all participating schools, Link Teachers from 75% (N = 170) of intervention schools were able to attend both training sessions as planned or via an online catch-up. Six schools changed Link Teacher during the programme, which mean that their Link Teacher only attended one of the sessions. However, there were also other schools where the Link Teacher only attended one session. Attendance at the first session was higher than attendance at the last (97% compared with 83%). Link teacher training attendance was also consistently higher in Cohort 1 than Cohort 2 for both training sessions, with just over two-thirds (69%) of Link Teachers in Cohort 2 attending the final training session either as planned or as an online catch-up compared with 86% of Link Teachers in Cohort 1 (see Figure 10).

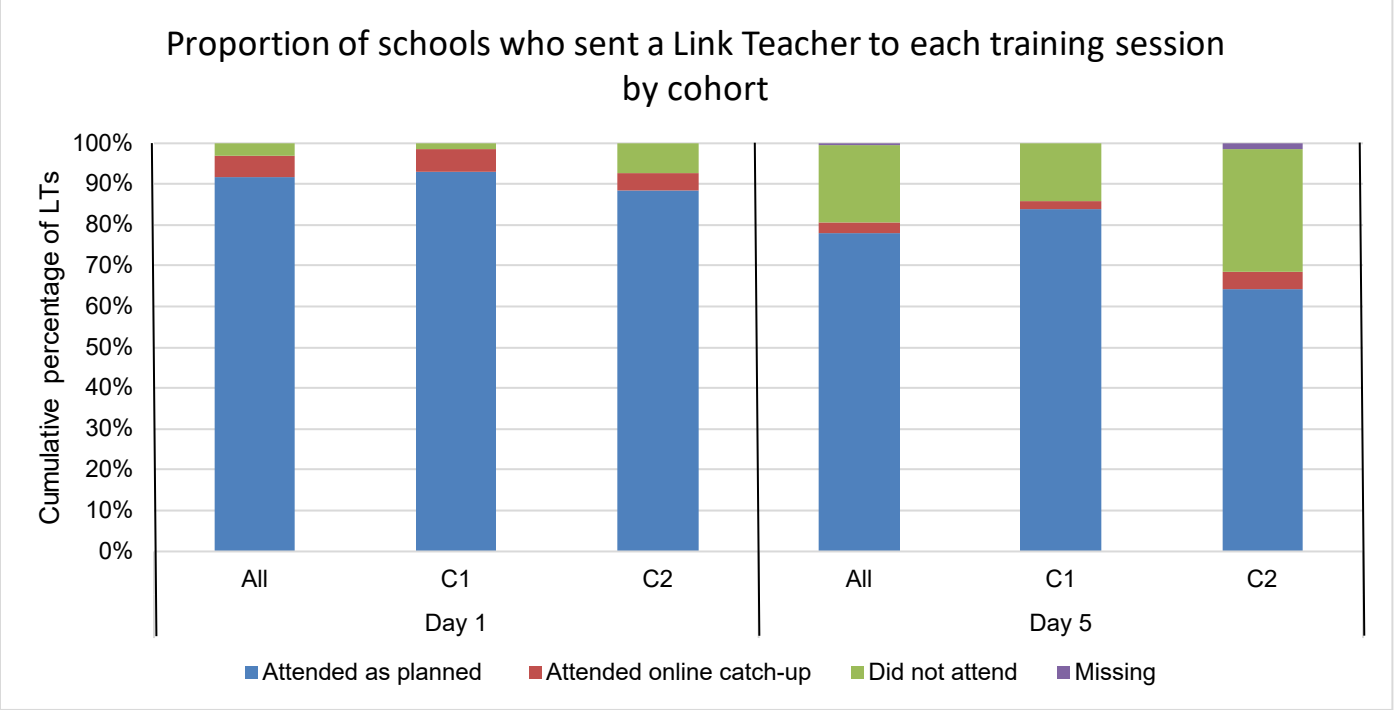
Figure 9: TA attendance at training sessions



Data from EHU on TA training session attendance. Percentage of schools that were able to send a TA to each training session as planned, attended an online catch-up session, or did not attend (all N = 226; C1: N = 156, C2: N = 70). Where schools changed TA part-way through the training, attendance data for both TAs has been combined.

[39] That is, one TA attended the first few training sessions then another took over delivery of the programme and attended subsequent training sessions but without having had the benefit of attending the early sessions.

Figure 10: Link teacher attendance at training sessions



Data from EHU on Link Teacher training session attendance. Percentage of schools that sent a Link Teacher to each training session as planned, attended an online catch-up session, or did not attend (all N = 226; C1 N = 156; C2 N = 70). Where schools changed Link Teacher part-way through the programme, data has been combined to show if the school had a Link Teacher present at each session, even if different people attended the sessions.

Pupil intervention attendance (RQ2.2, 2.3)[40]

Data from the pupil attendance registers shows that intervention group pupils received between zero and 30 intervention sessions (see Figure 11). Where pupils received no sessions (N = 22) this was mainly because they had left the school (N = 13) or withdrawn from the trial (N < 10). Overall, only 37% (N = 310/838) of pupils for whom we received attendance data received all 30 sessions (compliance measure 1). Compliance was higher in Cohort 2 than Cohort 1 with 41% of pupils in Cohort 2 (N = 102/248) receiving all 30 intervention sessions (compliance measure 1) compared with 35% of Cohort 1 pupils (N = 208/590). Compliance was also slightly higher among non-FSM pupils compared with FSM pupils with 39% of non-FSM pupils completing all 30 intervention sessions compared with 36% of FSM pupils.

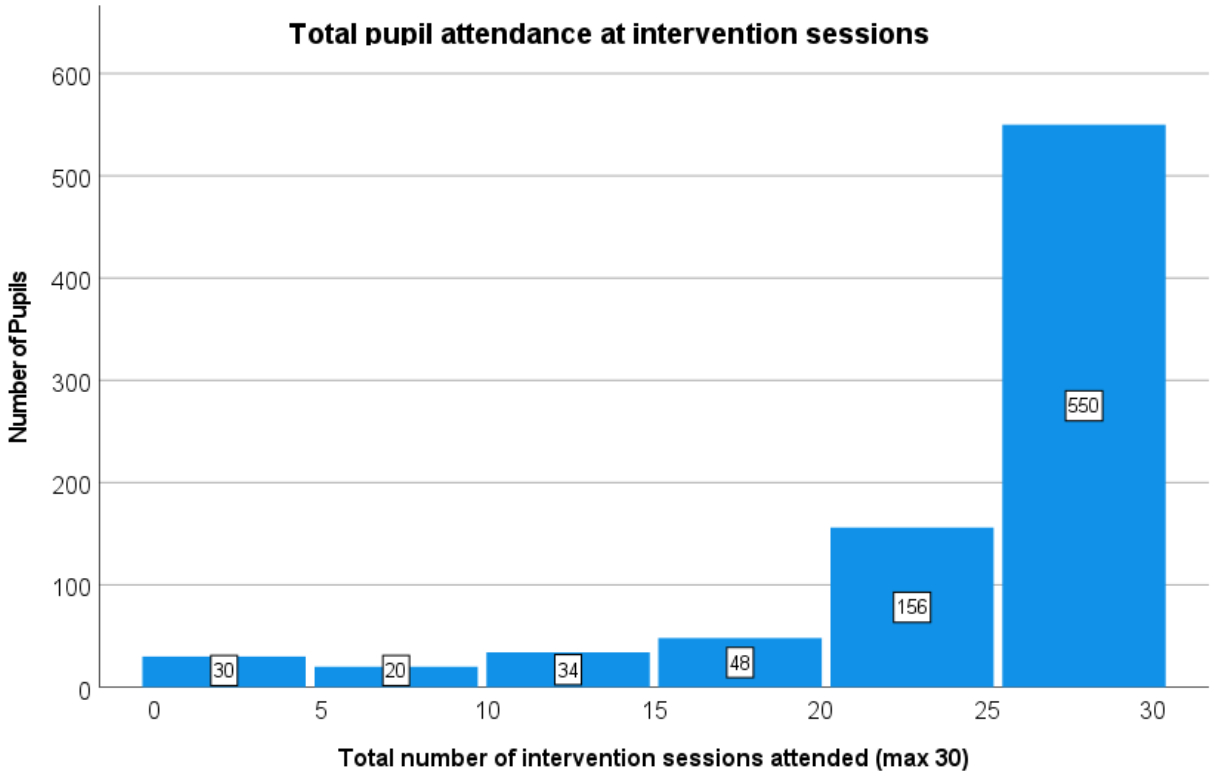
Just over half of pupils (56%) received at least 25 sessions in total including at least five out of six sessions for each topic (compliance measure 2). This time, compliance was slightly higher in Cohort 2 than Cohort 1: 58% of pupils (144/248) in Cohort 2 received at least 25 sessions in total including at least five out of six sessions for each topic (compliance measure 2) compared with 56% (328/590) in Cohort 1. Again, compliance was slightly higher among non-FSM (60%) compared with FSM (54%) pupils.

These figures suggest that, overall, full compliance with intervention delivery (all 30 sessions attended) was low in both cohorts. This is likely to have negatively impacted intervention effectiveness as more than half of pupils did not receive the full programme. These results contrast with case study interview data, which indicates that TAs were satisfied the majority of pupils had received the intervention in full. However, interviewees noted that there were cases where they had been unable to deliver the full intervention in advance of follow-up testing, or where a single pupil had been absent too regularly for TAs to be able to catch them up. One of the potential reasons for this discrepancy between the registers and qualitative results may be because a significant proportion of pupils missed two or more sessions from a single

[40] Note that all pupil-level data reported in the IPE section is for pupils for whom we received implementation data (for example, registers, maths intervention records). This differs from the impact evaluation which is based on pupils who had baseline and endpoint QRT data.

topic but otherwise had good attendance at the sessions. Close to three quarters of pupils (73%) received at least 25 sessions but missed two or more sessions from one topic. Pupils primarily missed intervention sessions due to school absence, so several days of absence close together (for example, a week off for illness) would explain why pupils missed multiple sessions from one topic but otherwise had good attendance.

Figure 11: Pupil attendance at intervention sessions



Data from pupil attendance registers. Columns represent 0–4 sessions, 5–9 sessions, and so forth.

Interview data from TAs indicates that where TAs were allocated time in addition to the intervention sessions (that is, the TA had the intervention timetabled over four or five days, rather than three) they were more able to provide catch-up sessions to pupils who had been absent. They also had the flexibility to pre-teach a session to pupils who they felt would struggle within the group, or to repeat a session for pupils who struggled during the main session. In comparison, those only allocated three delivery sessions per week found it much more challenging to offer catch-up sessions. Link Teachers also gave anecdotal evidence that absence rates amongst FSM pupils tended to be higher than for non-FSM pupils. This was supported by the intervention attendance register data (see above). This may have contributed to the relatively low level of compliance seen for the trial, given the prioritisation of FSM pupils, although we note the difference between FSM and non-FSM pupils was small.

Fidelity and adaptations (RQ2.4)

To ascertain fidelity, the research team carried out observations of all 13 trainers and observed programme delivery in ten schools. Observations were scheduled to allow the research team to observe a range of training and intervention sessions across each of the five topic areas. Through the survey and case study interviews, Link Teachers and TAs also provided details on any adaptations they had made to delivery. The findings from these observations, surveys, and interviews suggest that, overall, the training and intervention sessions were delivered with high fidelity to the recommended approach.

Training sessions—fidelity

The 1st Class@Number 1 training sessions are highly manualised. The delivery team reported that trainers are provided with presentation slides, detailed training notes, materials, and scripts to provide structure to the sessions. Scripts include content that trainers must talk about in their sessions, and ideas of content they could cover if necessary. The

delivery team reported that outside of the trial, development leads would conduct visits to trainers' sessions to monitor the consistency and quality of delivery. To reduce burden on trainers, this was not done during the trial due to observation visits from the NFER research team and delegates from the DfE and the EEF. However, the NFER observation visits did not serve a quality-assurance purpose, meaning the usual quality assurance of session delivery and support process from EHU was not present during implementation through the trial. These processes are further covered in the Scale-Up section of this report.

Through the reflective interviews, which followed the research team's observations, trainers reported that they were satisfied they had covered all the content they intended to in the session. They perceived their delivery to be similar to that outside of the trial. Researchers observed trainers providing Link Teachers and TAs with guidance on delivering the intervention with fidelity and the need for Link Teachers to advocate for the TA's delivery of the intervention. These elements were most strongly emphasised during the first training session but were reiterated over subsequent sessions. The researchers felt that, in general, trainers delivered engaging, interactive sessions in which TAs were able to discuss and share their experiences and gain ideas from other TA's practices. In addition, trainers modelled elements of the intervention as they would expect TAs to deliver, providing TAs with the chance to 'be the pupils', which supported their understanding of how pupils might approach the intervention. In the online session, this was executed by the trainer moving their screen so TAs could see them modelling the activities on their table. Trainers provided TAs with support and solutions for dealing with common misconceptions and pupil difficulties they were likely to encounter during the lessons. They also provided suggestions for how TAs might adapt the sessions in line with pupils' ability and group dynamics, which researchers observed TAs enacting in their lessons.

Training sessions—adaptations

The delivery team reported that they were unaware of any adaptations that trainers had made to delivery of the training through the trial. However, they acknowledged that there would be some differences based on trainers' prior experience of delivering and supporting the intervention, and it was felt that additions and advice based on these experiences were valuable for school staff. Aligned with this comment, the research team observed differences in the extent to which trainers drew upon their own experiences of delivering the programme and the tips they provided to TAs based on these experiences, which were not on the presentation slides. As delivery progressed, some trainers had also included additional slides, compared to what NFER were provided with, which were mostly used to facilitate discussions or showcase photos sent in by TAs (for example, showing their room set-up).

The most notable delivery difference the research team observed was the extent to which sessions included practical and interactive opportunities for TAs. Trainers' opinions over this generally fell into two groups. One group felt it was important to cover the content and theory through the training and that TAs could review the lesson plans and activities during their preparation time in school. In comparison, the other group of trainers acknowledged that the training forms an important part of TA's preparation, particularly for those who received little to no time to do this in school. This meant that having the chance to review and practice activities with other TAs was crucial for understanding how to deliver the intervention back in school. The online training also differed in this respect. The research team observed fewer practical opportunities and discussions among TAs. Whilst the trainer effectively modelled to TAs the activities on screen, they acknowledged that TAs may not have had the topic resources box and manipulatives (such as counters, Numicon, or coins) to hand, which made the online training more didactic than their in-person training delivery. The delivery teams' and school staffs' perceptions of these sessions are further discussed in the Responsiveness and Quality sections of this report.

Another key difference related to the REDS ('review, evaluate, discuss, share') sessions, which take place at the beginning of training sessions two to five and allow TAs to reflect on their delivery of 1stClass@Number 1, including what is working well and any challenges they have encountered. Given these sessions were based on TAs' experiences, they required careful trainer facilitation to ensure they were not dominated purely by negative comments related to delivery challenges and pupil groupings, which trainers were conscious might impact other TAs and set the tone for the remainder of the session. Trainers' approaches to facilitating these sessions generally fell into two groups: one group of trainers acknowledged these challenges but, overall, kept the session focused on positives in order to set the tone for the rest of the session; the other group of trainers allowed TAs more time, either in pairs or when feeding back to the group, to discuss the details of the challenges they were facing, which were often contextual rather than related to the intervention itself. In this latter group, the REDS session generally took longer, meaning there was less opportunity to cover the

handbook content within the session. This suggests that some further training for trainers may be beneficial to maximise the session time available for intervention content.

While all training sessions finished on time, the time spent on certain content or activities often differed from the timing on the presentations provided to NFER. The research team and trainers also identified challenges related to timing of the sessions. Trainers commented that being able to cover the REDS session, topic content, check understanding, and provide sufficient practical and discussion opportunities requires careful time management. This challenge was particularly salient in relation to the delivery of session three, which covers topics three and four. These two topics are covered together as they both focus on addition and subtraction. However, in two out of three of these sessions observed by the research team, much of the time was spent on REDS and topic three, with less opportunity (30 minutes compared to the intended one and a half hours) for TAs to review the topic four lessons and activities. This was consistent with the research team's observation that TAs felt less prepared to deliver the topic four lessons.

The researchers observed a small number of technical challenges during the training observations; however, these could not be avoided given trainers' reliance on venues' IT equipment. Where such glitches occurred, trainers proceeded with delivery using the handbook and their knowledge of the programme, which meant the IT difficulties did not detract from the training content.

Intervention sessions—fidelity

The 1stClass@Number 1 sessions are highly manualised, providing TAs with comprehensive lesson plans, scripted questions, and resources, thus supporting fidelity to the intended delivery approach. Qualitative data from the case studies and lesson observations suggested that the intervention was largely delivered with fidelity. TAs had the intervention timetabled three times a week, outside of pupils' usual maths lessons, as advised by the delivery team and trainers to ensure pupils did not miss whole-class maths teaching. However, the extent to which TAs could consistently deliver three times per week differed. Trainers advised that sessions should take between 30 and 40 minutes to deliver. It was noted from case studies and lesson observations that 40 minutes allowed TAs to provide a good balance between sufficiently covering all the lesson content without feeling pressured while retaining pupil concentration. For example, the research team observed one intervention session exceeding one hour, which the researcher noted to be detrimental to pupils' engagement and their understanding of the content, with pupils getting distracted and making more mistakes as the session progressed. In comparison, TAs who only received 30 minutes in which to deliver did not feel they could adequately cover each of the lesson elements or fully ensure pupil understanding of each concept covered. These experiences are discussed in the Facilitators and Barriers section of this report.

TAs provided with a separate room in which they could regularly deliver the intervention created permanent displays showcasing pupils' work and included resources, such as hundred squares and number tracks, for pupils to refer to during sessions. TAs working in schools with limited space achieved this by using portable whiteboards or flipchart books to display pupils' work.

Researchers noted during lesson observations that TAs were prepared for their lessons, having all the required resources and materials available and organised ahead of the lesson, which were then used as stipulated by the handbook. TAs consistently reinforced and emphasised the mathematical language as intended and encouraged pupils to use the correct language through effective questioning and modelling.

Intervention sessions—adaptations

As reported earlier in this section, the training provided TAs with a range of adaptations they could make to the intervention sessions. In addition, the topic handbooks provide TAs with suggestions for making each lesson easier and ways to extend pupils' learning. Most of the 174 Link Teachers who responded to the survey (91%) reported that their TA had adapted the 1stClass@Number 1 sessions to meet the needs of pupils in ways suggested by the trainer, while almost all of the 187 TAs themselves who responded to the survey reported making such adaptations (97% agreed or strongly agreed). However, when asked about specific delivery adaptations, 5% of TAs reported that they had not made any, suggesting that 2% of TAs answered these questions inconsistently. The adaptations reported, through the survey and the case study interviews, generally aligned with trainers' advice or the handbook, however, there were some exceptions. Both of these (aligned adaptations and further adaptations) are outlined below.

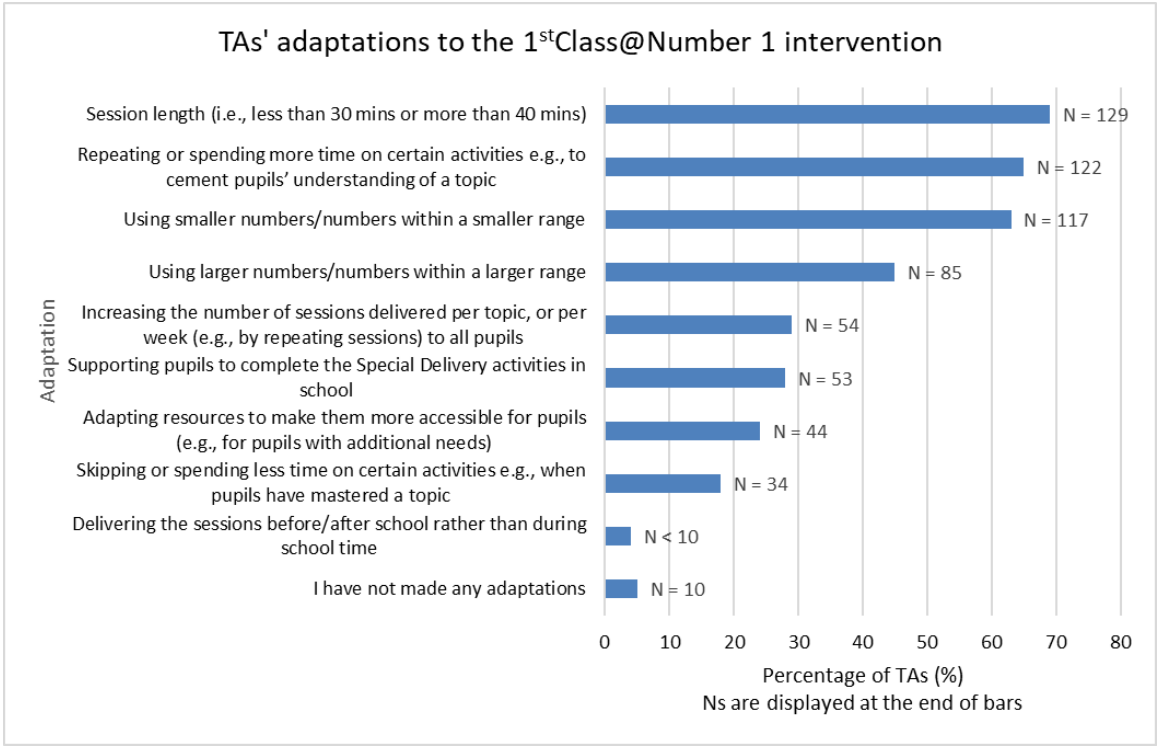
In the survey, the most commonly reported adaptation TAs made was in relation to session length (that is, less than 30 or more than 40 minutes—69% of the 187 respondents). While this adaptation was not noted by trainers or in the handbook, trainers acknowledged that early sessions may take longer as TAs gain familiarity with the intervention. Ability-based adaptations, aligned with guidance from trainers or the handbooks, were common: around three-fifths of TAs reported adaptations to better support pupils and around two-fifths reported adaptations to challenge pupils. Aligned with these findings, TAs interviewed in case study schools reported that their adaptations tended to be in response to pupil needs and abilities. This typically involved making the activities described in the handbook easier or harder by working with smaller or larger numbers or changing the range of numbers used, in line with the adaptation guidance. In some cases, it also involved simplifying or repeating the language more frequently to support EAL pupils, as advised by trainers.

Just less than a third of TAs reported through the survey that they had increased the number of sessions delivered per topic or per week to all pupils (29% of the 187 respondents). While this adaptation was not included in the handbooks, trainers suggested that if TAs had flexibility within their timetable, pupils may benefit from receiving certain lessons (on money and multiplication and division) over two sessions due to the complexity of the content. A small number of case study TAs reported following this guidance so they could cover the content at a slower pace. Around a quarter (24%) of TAs reported in the survey that they had adapted resources to make them more accessible for pupils. Again, while this adaptation was not listed in the handbook, trainers advised TAs to make SEND-based adaptions to resources as they usually would to support pupils in the classroom. Smaller proportions of surveyed TAs reported skipping or spending less time on certain activities (18%) or delivering the session before or after school—adaptations not suggested by trainers or the handbooks.

Just over a quarter (28%) of the 187 TAs who responded reported in the survey that they had supported pupils to complete the special delivery activities in schools. TAs who expected or found parent/carer engagement with the intervention to be poor reported, either through the case study interviews or during training sessions, that they had adapted how pupils engaged with special delivery elements of 1stClass@Number 1. Rather than sending the special delivery notes home, TAs supported pupils to deliver the notes to other members of staff in school and they kept the special delivery games in school for pupils to play at breaktimes with school staff or with older pupils.

Figure 12 illustrates the percentage of TAs who reported making the different types of adaptations to their delivery of 1stClass@Number 1.

Figure 12: TAs' adaptations to the 1stClass@Number 1 intervention



Data from the TA survey: 'Please can you tell us about what kinds of adaptations (if any) you made to the 1st Class@Number 1 intervention)', (N = 187). Items from the survey question have been reordered so adaptations made are presented in descending order.

Most TAs interviewed reported making no other adaptations to the intervention, although a small number reported minor adaptations to delivery which were not suggested during the training or in the handbook. One TA reported that they gave pupils number-writing tasks to complete while they waited for other pupils in the group to finish the intervention activities because they felt pupils needed additional support and practice with their number formation. Other TAs reported providing one resource per pupil rather than to pairs, or vice versa, based on pupils' needs. These adaptations were also observed by the research team during the lesson observations. Another TA reported that during a 1st Class@Number 1 lesson on number bonds, in addition to covering the intervention content, the pupils had used iPads to play a game usually played in the classroom to support their understanding of number bonds to ten. A small number of TAs reported, at the observed training sessions, lesson observations, and case study interviews, that, in line with guidance from their trainer, they had delivered the sessions in pairs rather than a group or four as pupils required more intensive support. This is further discussed in the Context section of this report.

During the lesson observations, researchers observed that TAs consistently followed the handbook. However, the extent to which they needed to keep referring back to the lesson plans reduced over the course of the intervention as they became more confident with their delivery and the structure of the lessons, which followed the same format each time. The research team observed few other adaptations to delivery during lesson observations, however, those that were noted were in relation to pupil ability (for example, providing numbers within a smaller or larger range) and session timing (as noted earlier in this section, one lesson observation took substantially longer than 40 minutes).

Quality (RQ2.4, 2.6)

Observations of the training sessions and delivery of the intervention in schools were used to determine the quality of 1st Class@Number 1 delivery. Through the survey and case study interviews, Link Teachers also spoke of their perceptions of the quality of TAs' delivery. Overall, the research team perceived the intervention to be of high quality, which was facilitated through trainers' and TAs' level of experience and enthusiasm for the programme.

Quality of the training sessions

Data from the survey demonstrates that Link Teachers and TAs perceived the 1st Class@Number 1 training sessions to be of high quality. Almost all respondents agreed or strongly agreed that the training content, delivery, and materials were of a high quality (96% to 98% of the 174 Link Teachers and 187 TAs across these three areas of the programme). Similarly, high proportions agreed or strongly agreed (95% to 97% of Link Teachers and TAs) that they were satisfied with the trainers' knowledge of the intervention and how to implement it. They felt that the training sessions were engaging, with sufficient opportunities for discussions and to practice activities (95% to 98% of Link Teachers and TAs agreed or strongly agreed). The vast majority of TAs and Link Teachers (93% and 98% respectively) also found the support provided when the trainer visited their school helpful. Despite not conducting their own observations during the trial, the delivery team reported that they were satisfied the in-person training sessions had been delivered with the usual high quality they would expect.

Based on their observations of the training sessions, the research team, likewise, perceived the training to be of a high quality. Researchers observed that the trainers demonstrated passion for, and knowledge of, the 1st Class@Number 1 intervention. As noted in the Fidelity and Adaptations section of this report, most trainers delivered engaging, interactive sessions in which they shared their practical experiences of the programme. This appeared to be well-received by Link Teachers and TAs and facilitated engagement as it brought the intervention to life. Trainers were also able to tailor aspects of their delivery to ensure it met TAs' needs based on their knowledge of the contexts in which TAs were working which was gained through school visits and/or prior relationships with schools. Trainers ensured contributions to discussions came from across their groups and acknowledged contributions with gratitude and encouragement. This in turn created a supportive, collaborative learning environment for the TAs. Trainers checked TAs' understanding of the training content and modelled how TAs should deliver the intervention to pupils—a model the research team perceived to be valuable for ensuring TAs understood how to deliver the intervention with fidelity.

However, there was some indication from across the evaluation activities that the online training sessions may have offered a different experience to the in-person training sessions. Across survey items related to delivery quality and interactivity, levels of agreement, though still high, were lower than for the in-person training groups. For example, TAs

less strongly agreed that the online training offered sufficient opportunities to practice activities (online group TAs: 25% strongly agreed and 58% agreed; in-person group TAs: 70% strongly agreed and 26% agreed), and that training sessions were engaging (online group TAs : 42% strongly agreed and 50% agreed; in-person group TAs: 72% strongly agreed and 25% agreed), with a similar pattern of results seen for Link Teachers. These results align with the research team's observations of lower engagement and fewer interactive opportunities in the online compared to in-person sessions (see Fidelity and Adaptations). They are also consistent with views from the delivery team and the trainer, who felt that it can be more difficult (though it is possible) to build relationships online and noted that the interactivity of sessions is dependent on TAs bringing their resources to the training. It was also noted that the trainer had limited access to the functionalities available on MS Teams (for example, breakout rooms, chat function), which is something that would need to be addressed if online training is to be used more widely. It is important to recognise that these results come from one small group of participants and the observation of one online training session, but they suggest that potentially some further development work is needed before the online training could be rolled out more widely.

Quality of intervention sessions

Just over half of the 174 Link Teachers who responded (53%) reported in the survey that they had observed their TA's delivery of 1stClass@Number 1. However, almost all Link Teachers (98%) reported that they were satisfied with the quality of TAs' overall delivery of the intervention (irrespective of whether they had observed delivery). Similarly, case study Link Teachers' capacity to observe their TA's delivery of the intervention was mixed. Link Teachers who had conducted an observation were happy with the quality of TA's delivery. Those who did not carry out an observation said this was due to the intervention being delivered while they were teaching their own class. However, they reported that their TA had 'just been able to get on with it'.

The research team also perceived TAs' delivery of the intervention to be of a high quality, including where the TA had been trained online. They observed that TAs delivered the intervention with confidence, particularly as the trial progressed. They generally delivered smooth, pacy sessions which kept pupils engaged that, where necessary, were adapted to suits pupils' need and abilities. TAs used their handbooks to guide their delivery, which supported them in consistently questioning pupils, including using 'trick' questions to check for misconceptions and modelling the mathematical language which pupils emulated.

Responsiveness (RQ2.5, 4.6)

As part of the Responsiveness dimension of the IPE, the research team sought to understand how Link Teachers and TAs responded to, and engaged with, the 1stClass@Number 1 training sessions. The research team also explored how pupils and parents/carers engaged with the intervention. Data from the training and lesson observations and case studies suggest that overall, the 1stClass@Number 1 intervention was well received by school staff and pupils, but parent/carer engagement was low.

TAs' and Link Teachers' responses to the 1st class@Number 1 intervention

The research team perceived TA and Link Teacher engagement with the training sessions to be high. They observed TAs and Link Teachers taking notes on and discussing the intervention activities, working in pairs and sharing their experiences of delivering the intervention with the wider group. As discussed earlier, engagement of TAs in the online group was more difficult to gauge but appeared to be lower compared to in-person training, both within the main session and in breakout rooms (which were designed to replicate the group discussion opportunities provided during in-person training sessions). The research team identified that, overall, online discussions were shorter and less varied due to the same few TAs contributing and sharing their experiences. As acknowledged by the trainer and discussed in the Fidelity and Adaptations section of this report, there were also fewer practical opportunities due to the nature of online training in which TAs were not sat together (making paired or group practical activities impossible). Within each topic, the handbook listed resources that schools would be required to provide. Trainers of in-person sessions brought these resources along to sessions for TAs to use, however TAs attending online may not have had these resources to hand, meaning these sessions were perceived to be less interactive than in-person delivery. This may be remedied in any future online training sessions by reminding TAs in advance of the sessions to bring along additionally required resources that are available in school, as well as their topic resource boxes.

Link Teachers and TAs in case study schools reported that they had been satisfied with the training content, which they described as 'detailed and 'informative'. Link Teachers felt the training model, in which they only attend days one and

five, worked well. They found it valuable to gain details of how to support the intervention in the first session, and how to make the intervention sustainable in the final session but acknowledged they did not need to be trained in the detail of each intervention topic. TAs also liked the incremental nature of the training, which was preferred to receiving intensive training of all five topics at the start of delivery, as had been experienced with other interventions. This was seen to support their delivery because content from the training was fresh in their minds when it came to delivering the topic in schools. Link Teachers and TAs interviewed perceived the training delivery to be of a high quality. They appreciated trainers' enthusiasm, experience, and knowledge of the programme which enabled them to provide TAs with a range of ideas and solutions to support delivery. There was equally high satisfaction with the quality and provision of resources to deliver the intervention. Interviewees found it particularly helpful that they could download additional copies from an online portal. Link Teachers and TAs commented that these features made 1stClass@Number 1 stand out from other interventions they had delivered, which schools often had to resource themselves. TAs reported that the provision of resources meant they had everything they needed, which saved them time when preparing to deliver the intervention and said that any additional manipulatives required could easily be found in school.

> *'The other thing we found really good was that the resources were provided. That is a big thing in school, not having the money, the time as well, but everything was provided…The resources were really good and we have been able to download them, so I have been able to use extra resources, send them home which was really good' (TA, case study school).*

Following the first training session, all case study Link Teachers and TAs received a visit from their trainer, either in-person or virtually. Case study interviewees found this visit helpful for ensuring they had everything in place to begin delivering the intervention—a room to deliver, sufficient preparation, and delivery time for the TA—and to ask any questions which arose from early delivery. There were very few suggested improvements to the training and support provided as case study participants generally felt they had everything they needed, but one suggestion made was that it would have been helpful to have an additional virtual trainer visit mid-way through the intervention to check in on progress.

TAs reported that they had enjoyed delivering the intervention. In the survey, 97% of the 174 Link Teachers and 98% of the 187 TAs who responded reported this to be the case and a similarly positive response was achieved in the case study schools in relation to TA enjoyment. There was also consensus across the case study Link Teachers that TAs had shown commitment to, and enthusiasm for, the intervention.

Trainers' perceptions of TA and Link Teacher responses to the intervention

Trainers reported that they had been satisfied with the level of engagement they achieved at the training sessions the research team observed. Trainers perceived the delivery of day one to be the most didactic session of the five days as it covers a range of content related to theory, prior evidence of impact and guidance for successful delivery. Trainers commented that TAs were appreciative of the training and were keen to take responsibility for delivering an intervention. Trainers also reported that TAs' engagement from day two of the training onwards improved further as they became more familiar with the intervention and group relationships started to build. This was also reflected by case study TAs who spoke of the supportive group dynamic at the training sessions and appreciated the opportunity to share experiences, which they found beneficial for informing their own delivery. Trainers felt that the rapport they had built with TAs and the supportive relationships within the group had contributed to TAs' willingness to engage with group discussions.

TAs' and Link Teachers' knowledge, skills, and confidence to deliver the training

Most of the 187 TAs (98%) and 174 Link Teachers (94%) who completed the surveys felt that the training fully prepared them for their roles delivering and supporting the intervention. Similarly, case study TAs and Link Teachers reported feeling prepared to deliver the intervention in school, which they attributed to the high quality, comprehensive training delivery and provision of intervention resources.

Overall, trainers were satisfied that Link Teachers and TAs were prepared to deliver each of the 1stClass@Number 1 topics following the sessions observed by the research team. However, under the EHU delivery model there is flexibility for trainers to focus on the elements that they feel are most important. This led to differences in delivery approaches among trainers and consequently they attributed TAs preparedness to deliver the sessions to different things. Trainers who had provided lots of opportunities for TAs to review the lessons and who gave TAs the chance to practice the

activities with each other during the training attributed this to level of preparedness. Other trainers, who chose not to cover all of the lessons in detail but picked out what they identified to be the 'important elements', such as the theory, misconceptions pupils may encounter, and activities which TAs may struggle to interpret and deliver accurately, felt that TAs would be prepared but they would need to review the lessons more thoroughly in their preparation time. The research team noted comments from TAs in these groups that they would have liked more time focusing on delivery of the lessons and practicing activities and less reflection time. The two quotes below highlight these differences in trainers' approach, and their views of TA preparedness:

> *'I think they are super prepared! The training session forms a huge part of their preparation. They are leaving here today having read through everything, been able to do the activities for each of the lessons so they just need to have a quick scan over it before they deliver the lessons' (trainer, Cohort 1).*

> *'I think they are ready [to deliver]. A big part of them running it is to just make a start, if they can do this confidently, they'll pick it up. We can't cover everything, go through everything in every lesson they need to do, it would be cognitive overload. So, I picked the things to go into detail, the things they might find more difficult getting their heads around on their own' (trainer, Cohort 2).*

The handbook provides TAs with detailed lesson plans and scripted questions which trainers encouraged TAs to follow. Trainers reported that their experience delivering the programme allowed them to predict the misconceptions that TAs are likely to encounter, enabling them to provide TAs with solutions or adaptations to support pupils with these in advance. Trainers interviewed as the intervention progressed felt that TAs became more prepared and confident to deliver as they worked through the intervention because they become familiar with the structure of the topics and can anticipate what skills and understanding pupils will need to have secured in early lessons to progress.

There were concerns around Cohort 1 TAs' preparedness to deliver topics three and four due to the timing of this session in the school year. This session was delivered in December and the extent to which Christmas activities disrupted delivery of the intervention was mixed across schools. Ideally, TAs would have completed topic two ahead of this training session but this was not always the case. Trainers acknowledged that TAs who experienced significant disruption would not deliver topic three and four until January, which would be over one month after receiving the training. Trainers hoped that TA's effective note-taking during the session would help reduce the impact of this lag time between receiving the training and delivering the topics. The research team observed the delivery of one topic three and one topic four lesson in Cohort 1. One of these TAs did report that they had found the delivery of these topics difficult given the time between training and intervention delivery but had found their notes helpful. In both cases, the research team observed TAs referring back to the notes they had made during the training session but, overall, perceived TAs to have delivered effective sessions with confidence. We also note that in usual implementation of 1st Class@Number 1, training and delivery usually start earlier in the autumn term and finish before Christmas. Some of the concerns with the timing of sessions noted by TAs related to training and delivery starting later in the autumn term after the collection of baseline data and randomisation.

Pupils' response to, and engagement with, the intervention

Evidence from the surveys, lesson observations, and case study interviews with school staff and pupil focus groups indicated that pupil response to 1st Class@Number 1 was very positive. Almost all of the 174 Link Teachers (99%) and 187 TAs (98%) who completed their respective surveys agreed or strongly agreed that pupils had enjoyed the intervention. Given that almost all pupils appeared to enjoy the intervention, it seems unlikely that differences in intervention enjoyment drive differences in intervention outcomes. A slightly lower percentage, but still the vast majority, reported that pupils in the intervention group had worked well together (88% of Link Teachers, 80% of TAs).

The qualitative data activities provided further insights into pupils' engagement with the intervention: there was consensus across case study Link Teachers and TAs that pupil engagement with the intervention was high. They commented that pupils had shown enthusiasm for attending their maths group and had enjoyed participating in the games and activities. TAs also recalled that pupils were disappointed to hear in the last session that they had completed the programme. Case study staff felt that high pupil engagement had been facilitated by the fun, practical nature of the intervention which they felt often contrasted to pupils' usual maths lessons – a point also emphasised by trainers.

> *"The kids have absolutely loved it. I think because it's with games, learning through play, the amount of progress they have made is unbelievable, they have really enjoyed it. They even asked if we could play*

some of those games during golden time so when they are choosing and wanting to show their friends[41] what they have been doing. They have loved it. It was interesting at the end[42] because I thought they would all pick the same game, but they didn't, they all liked different games, so it was really nice." (TA, case study school)

During their lesson observations, the research team also noted that engagement from most pupils was high. They were observed paying attention to the TAs, joining in the activities, and asking and answering questions during the sessions. However, the research team noted that in the lesson observation in a school from the online training group, which exceeded 40 minutes, pupil engagement decreased in the latter parts of this session. Despite this, they noted that for the first half of this session, and in all other sessions, pupils were happy to attend the intervention, enjoyed the activities, worked well in pairs and as a group (discussed further in Facilitators and Barriers), and responded well to the TAs' questioning.

Pupils' thoughts on the intervention

Through the pupil focus groups, pupils in case study schools fed back about what they had liked and disliked about 1stClass@Number 1 as well as what they felt could improve the intervention. It should be noted that pupils found it easier to identify what they had liked rather than what they had not liked or would change (which can be expected given the level of reflection required for this question and the pupils' young age of six or seven). Their views also differed based on individual preference.

Pupils liked the activities and games. Those mentioned most often included the teacup game (in which pupils distinguish 'teen' and 'ty' numbers),[43] games using the parcels and delivery vans, the race game, one more, one less game, and painting their hands to create a five number track.

'I liked counting in twos, fives, and tens. It made it easier when doing our times tables. I also liked one more one less game' (Pupil, case study school).

Pupils also said they liked the location of the sessions (such as in a 'special' intervention room), working in pairs or with their TA, decorating their exercise books, and taking postcards back to class or home to share their achievements.

'I got a postcard to take home, and my mummy was so happy, she said it was so good' (pupil, case study school).

Pupils reported finding some of the activities and games more difficult, or not as fun. This included finding missing numbers on a 100 square, counting money, and writing number sentences. However, there was no consistency in the games or activities pupils disliked and these tended to be individual preferences. Pupils' views were split about the timing of the sessions, which tended to be in the afternoon: while one group did not mind when the sessions took place, another did not like that they missed lessons such as music and PE. Pupils who were unable to play the special delivery activities at home would have liked to be able to do this. Pupils who TAs identified to the researcher as being more able pupils suggested that 1stClass@Number 1 could be improved by making it more challenging, such as working with larger numbers and counting in multiples of threes, fours, and 12s (in addition to twos, fives and tens).

'Counting in twos, I didn't like that because it was easy. I would change it to counting in threes. And counting in sevens instead of ones because that's tricky and we like tricky things because when it's tricky we learn' (pupil, Cohort 1 case study school).

In comparison, there were comments from pupils who TAs identified to the researcher as being of lower ability about making the intervention easier through working with smaller numbers in games. Aligned with comments from TAs, pupils reported disappointment that they could not continue with the intervention after they had completed all sessions. They

[41] We think that any potential contamination from this would be low as the intervention group children's friends would not have received the key elements of the 1stClass@Number 1 intervention.

[42] In the final session, there is the chance for children to play their favourite game or activity from the intervention again.

[43] 'Teen' numbers span thir**teen** to nine**teen**; 'ty' numbers are multiples of ten: thir**ty**, for**ty**, fif**ty** etc.

also reported that they would have liked more time and more opportunities to play the games again, providing a further indication that they had enjoyed the intervention.

> *'I want to do the handprints again but do it with twos and tens' (pupil, case study school).*

> *'I'd like to do it [the intervention] again at the junior school [next academic year]' (pupil, case study school).*

What were the challenges with pupil engagement?

A small number of challenges with pupil engagement were identified by the research team and the TAs who participated in the lesson observations. These were related to behaviour and SEND. Behavioural challenges included one or two pupils in the group disrupting others and some having difficulties sharing, working in pairs, listening to others, or allowing others the chance to answer. TAs found that for pupils with SEND, problems related to attention and retention difficulties and there were some issues with remaining on task, which could distract the rest of the group. In line with the research teams' session observations, TAs did not report any differences in engagement between FSM and non-FSM pupils, instead perceiving these differences to be based on pupils' ability and confidence.

Parent/carer engagement with the intervention

In comparison to the high pupil engagement reported, Link Teachers and TAs perceived that parent/carer engagement with 1st Class@Number 1 tended to be poor or that they did not know whether parents/carers had engaged. Around half of survey respondents reported that parents/carers had not supported their child to complete the special delivery activities at home (48% of the 174 Link Teachers and 54% of the 187 TAs); only around a third reported that parents had engaged with the special activities (37% of Link Teachers and 30% of TAs). The remaining proportion reported uncertainty around parent/carer engagement.

Data from the case studies aligned with these survey findings. Link Teachers and TAs reported that, overall, parent/carer engagement with the intervention tended to be low. TAs reported that pupils' response to taking the special delivery games home was mixed—they recalled that some pupils had been excited to do so, however, they were uncertain about the extent to which these had been played with parents/carers or siblings. Other pupils, who expected that games would not be played, or would be damaged, did not wish to take the special delivery activities home. In cases where parents/carers did engage with the activities, TAs reported that parents/carers had shared photos or asked when the next game would be sent home. Language barriers were identified by case study interviewees, trainers, and TAs during training observations as a challenge for engagement of parents/carers who spoke limited or no English. Trainers and TAs discussed the potential of TAs explaining the special delivery notes and games to older siblings who would be able to translate for their parents/carers or play the games at home with intervention pupils. A small number of TAs reported success with this approach to overcome EAL barriers. More widely, Link Teachers and TAs commented that the lack of engagement with the intervention by pupils' parents/carers reflected poor engagement with school activities—such as homework or home reading—more widely. This meant the lack of engagement with 1st Class@Number 1 was not a surprise but was still disappointing.

To further understand pupil and parent/carer engagement with the special delivery activities at home, pupils who participated in the focus groups had the chance to design a postcard illustrating them playing their favourite game at home. The most popular special delivery game was the race game (N = 16), followed by the teacup game (N = 10), number cards (N = 6), and 'ten nice things' (N = 5).[44] Pupils' reports on if and how they had engaged with these games at home differed. Pupils who had engaged at home reported that they had played the games with a parent/carer, grandparent, or sibling. They drew or wrote about their experiences of playing these games on their postcards. Pupils who had not taken or played the games at home still drew their favourite game but described themselves playing it at school, such as during the intervention sessions or with the TA during golden time or breaktime.

[44] Figures here do not sum to 43 (total number of focus group pupils reported elsewhere) as in one school, pupils did not have time to design a postcard and a small number of pupils in other schools who had not completed special delivery activities at home instead designed a postcard with number facts, rather than an illustration of a game.

Facilitators and Barriers (RQ3.1, 3.2, 3.3)

RQ3: How effectively was the intervention implemented in schools, and what supported or hindered successful implementation?

The logic model identifies a range of potential factors that may moderate the intervention outcomes. These include:

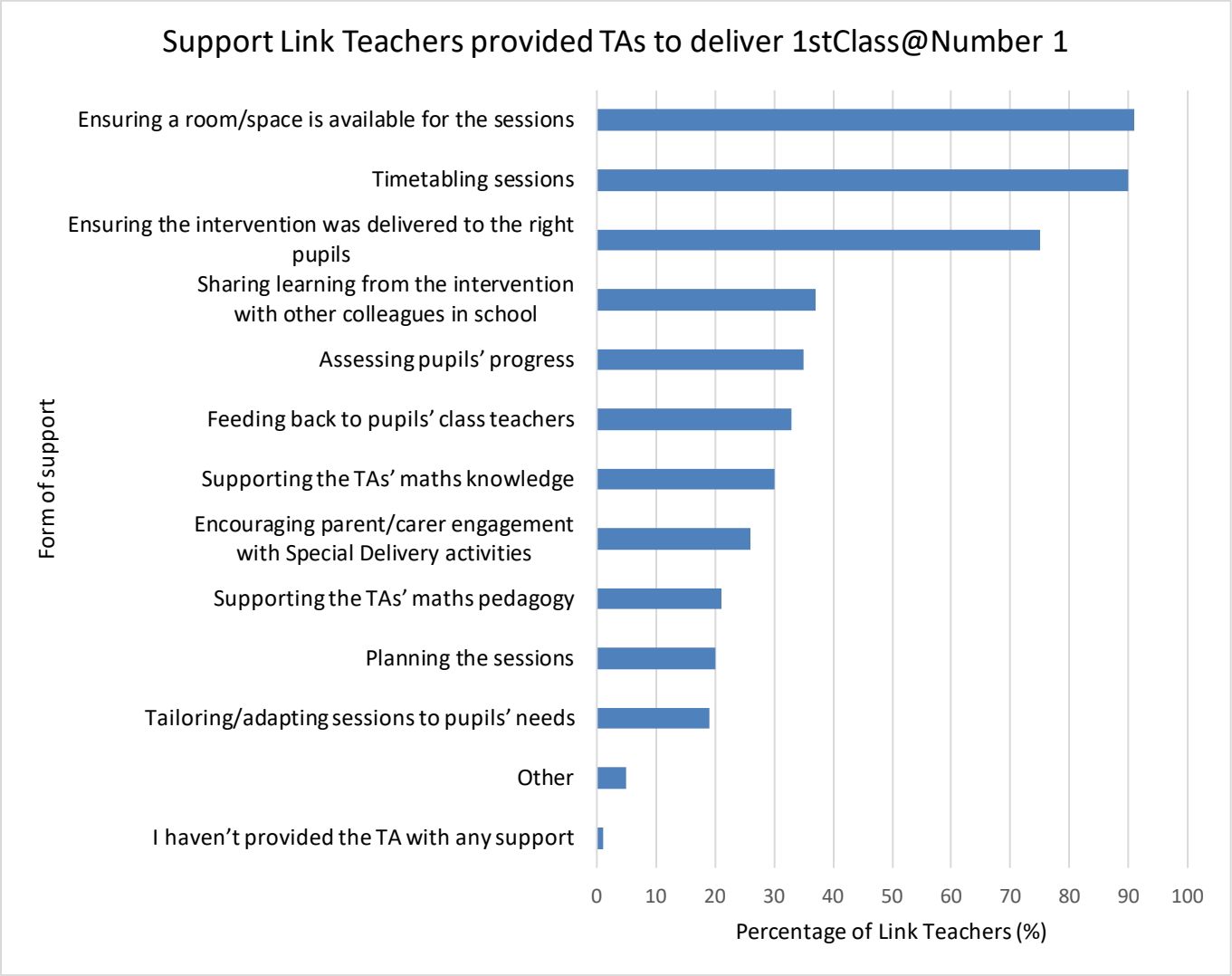
- characteristics related to the TA, such as their experience, capabilities, and intervention preparation time;
- characteristics related to the Link Teacher, including their management and coordination skills and time;
- characteristics related to pupils, including their fit with selection guidance, prior attainment, SEND, and English as an additional language (EAL);
- parental attitudes, confidence in, and knowledge of maths; and
- school-level factors including senior leadership and class teacher buy-in and engagement, the facilities available for intervention delivery, and the alignment of the intervention with the maths curriculum.

The research team explored these potential moderators and the extent to which they facilitated or presented a barrier to implementation through the case study interviews with school staff. The TA and Link Teacher surveys as well as comments made by trainers and TAs during training and lesson observations have also fed into this section. It is important to note that the barriers to delivery tended to be reported during observations whereas there were only isolated incidences within the case studies where TAs raised barriers. Additional factors supporting the implementation of the intervention also emerged from these conversations.

Link teacher support

The 1st Class@Number 1 logic model identifies Link Teacher support for the TA as a moderator for the intervention's success. The importance of this supportive role was emphasised to Link Teachers at the first training session. Most of the 187 TAs who responded (91%) reported through the survey that they felt supported by their Link Teacher to implement 1st Class@Number 1 and had received all the support they needed to deliver the intervention effectively. Through the survey, Link Teachers indicated the forms of support they had provided to their TA throughout delivery of the intervention (Figure 13). This support most commonly included ensuring a room or space was available for the sessions (91%), support to timetable sessions (90%), and ensuring the correct pupils received the intervention (75%) (see Figure 12 for full results). Aligned with these findings, case study TAs and Link Teachers reported that Link Teacher support tended to be in relation to ensuring the TA received sufficient time and space to deliver 1st Class@Number 1. TAs reported that being able to deliver the intervention in a quiet space, such as an intervention room, supported their delivery. It meant that pupils would not be distracted by the activities of other staff and pupils and allowed TAs to set up a learning wall to display pupils' work and resources that they could refer to during the sessions, which is something which is encouraged during the training sessions. Note that as this was a pupil-randomised trial, the guidance provided to schools as part of the evaluation asks that displays are not left on show if other staff or pupils outside of the intervention group can access the room the sessions are being delivered in. Timetabling support is discussed further in the section below.

Figure 13: Support provided by Link Teachers to TAs to deliver 1stClass@Number 1



Data from Link Teacher survey: 'What kinds of support have you provided to the TA delivering the 1stClass@Number 1 intervention?' (N = 172). Items from the survey question have been reordered to be presented in descending order.

Delivery and preparation time

TA's preparation time for the intervention is included in the logic model as a moderator for the intervention outcomes and the importance of TAs being allocated regular, sufficient preparation and delivery time was emphasised by the trainers. Link Teachers were crucial in securing TAs this time on their timetables. There were a range of approaches to timetabling which supported TAs. In cases where the intervention was timetabled for more than three times a week, this gave TAs sufficient time to prepare the sessions on days where they were not delivering the intervention. It also gave them the flexibility to revisit content with pupils who were absent or required repetition of the content to ensure they fully grasped it. In other cases, TAs received one hour on their timetables, which allowed them 20 minutes preparation time and 40 minutes delivery time. Trainers suggested that 40 minutes delivery time was optimum for covering the content in the right level of detail for pupil abilities.

In comparison, TAs not allocated preparation time included in their timetable reported needing to prepare at home or during breaks. Those who only had 30 minutes timetabled to deliver the intervention often found this was insufficient to deliver all the content detailed in the lesson plans. Researchers also noted following some 30-minute lesson observations, across the topics, that pupils would have benefitted from additional time to fully grasp the content, which TAs acknowledged. However, several case study Link Teachers commented that 40 minutes or more is a long time for TAs to be out of the classroom, when their support is so valuable to class teachers. Where TAs had just three sessions timetabled per week, they experienced difficulties catching up pupils who had been absent. Some TAs would have also

liked additional time to repeat or split up lessons in topic five—covering multiplication and division (an adaptation in line with guidance from trainers)—to secure pupils' understanding but not all had time for this ahead of the endpoint testing.

At the training and lesson observations, several TAs raised experiences of other priorities taking them away from preparing or delivering 1st Class@Number 1 such as other interventions (phonics is a competing priority in Key Stage 1), covering teachers, or whole-class or whole-school activities. As noted in the Responsiveness section of this report, the extent to which Christmas activities were felt to have disrupted delivery of the intervention was mixed across Cohort 1 schools. TAs who faced difficulties with continuing to regularly deliver the intervention around nativity and other Christmas activities faced disruption to the delivery of topics two and three. Additionally, TAs reported at the training that pupil absences increased at this time, which further delayed their delivery.

Pupil needs and group dynamics

Case study TAs reported that positive pupil engagement, behaviour, and their ability to work well together had facilitated delivery and TAs were particularly grateful for this during their first experience of delivering the intervention. However, researchers noted several comments during training and lesson observations of TAs who experienced challenges in delivering the intervention driven by pupil needs, poor behaviour, and difficulties managing the group dynamics.

TAs commented that pupils with SEND had difficulties retaining information, remaining on task, and recognising and transposing numbers. This meant these pupils required repetition of content, for which there was not always time. Challenges were also raised in relation to EAL pupils: for example, one trainer commented that across their group, many EAL pupils were receiving the intervention but that these pupils would have benefitted more from English language interventions as they were having difficulties accessing the 1st Class@Number 1 content. TAs also reported that some EAL pupils had difficulties with understanding and using the maths language to answer questions in full sentences and required simplification and repetition of this mathematical vocabulary. TAs working with pupils who used Arabic at home reported additional challenges: when TAs spoke about addition, or 'counting on' using a number line, in English, conceptually this moves left to right; however, this could be counter-intuitive to pupils familiar with reading and writing right to left, so additional time was required to allow for this code switching.

Due to the challenges with pupil selection and groupings highlighted above, there were cases where TAs expectations of what pupils could achieve were low. Trainers worked to dispel this with comments such as 'let them give it a go', 'you might be surprised'. For TAs with intervention groups comprised of pupils with varying abilities, their approach and ability to managing this effectively differed. Some reported at their training sessions that they found this difficult to manage because the content needed to be pitched differently across the group; other TAs used this to their advantage, for example, through pairing a stronger with a weaker pupil so the stronger pupil could scaffold and support the TA in their explanations to pupils who experienced difficulties with the content.

Researchers noted several comments during training observations from TAs who were unsatisfied with the randomisation, which they felt had led to difficulties in managing their intervention group. This dissatisfaction was generally due to disruptive pupil behaviour or notable variation in pupils' needs and abilities. It should be noted that pupil group selection is identified within the logic model as a moderator of the intervention outcomes. It is also important to note that these difficulties were due to the trial (rather than the intervention itself) and may in part be due to pupils unsuitable for the intervention being initially screened using the SENT-R and included in the randomisation as there was no lower cut-off score. Guidance for nominating pupils for initial screening advised schools to choose pupils who were working around 12-months behind age related expectations towards the end of Year 1 and were encouraged to nominate FSM-eligible pupils who met this criterion. NFER selected eight pupils for the trial based on SENT-R score and FSM-eligibility. Link Teachers could choose to switch out non-FSM pupils if they felt another pupil would be more suitable. Although the switching out of FSM pupils was not encouraged, teachers could make the decision to switch them out for other FSM or non-FSM pupils whom they felt were in greater need of the intervention. Despite these screening and selection procedures, some TAs still felt that there were other pupils in the class who they felt would have benefitted more because they believed pupils selected were too low or high ability for the intervention or the pupils selected for the intervention did not work well together or had behavioural challenges, which caused disruption and took up valuable time.

Support from other staff in school

The logic model identified senior leadership and class teacher buy-in and engagement as further moderators for the intervention outcomes. Most of the 187 TAs who completed the survey reported that they felt supported by senior leaders (88%), although this was higher among Cohort 1 TAs (91%) compared with Cohort 2 TAs (81%). Most also felt supported by the intervention pupils' class teacher(s) (93%) to implement the 1st Class@Number 1 intervention. Both Link Teachers and TAs felt that senior leaders had been supportive of the intervention taking place. Year 2 class teachers (who were not also the Link Teacher) were more involved directly with the intervention than senior leaders. TAs reported that Year 2 teachers had worked with them to identify suitable times for them to take pupils out of class. TAs also appreciated class teachers' engagement with the special delivery notes, such as reading these out and displaying them in the classroom. Generally, case study TAs were also the Year 2 TA, however, in cases where the TA delivering did not work in the pupil's maths class and where there was time to do so, Year 2 teachers and TAs had met to discuss the content of maths lessons and the intervention to ensure links could be made across the two to support pupils' learning and to discuss progress made.

Alignment of 1st class@Number 1 with the maths curriculum

There were high levels of agreement in both the Link Teacher and TA surveys that 1st Class@Number 1 aligned well with the school curriculum. Link Teachers were more likely to report alignment than TAs (94% and 87% respectively), potentially due to increased curriculum knowledge. By cohort, the proportion of Link Teachers reporting alignment was comparable (94% in Cohort 1; 92% in Cohort 2), however Cohort 1 TAs were more likely to report curriculum alignment compared to TAs in Cohort 2 (91% and 79% respectively). Link Teachers and TAs who were interviewed reflected on the alignment of 1st Class@Number 1 to the curriculum. They reported that the content of the sessions often aligned with the topics covered in class, which meant pupils could draw upon their learning from the intervention to support them in maths lessons or vice-versa. However, these comments were more frequently made by Cohort 1 staff and as demonstrated by the survey findings, there were instances of Cohort 2 staff reporting that the topics would have aligned better had they started the intervention earlier in the academic year due to the timing of topic delivery in the Year 2 maths curriculum. This suggests that alignment with the maths curriculum may be an outcome moderator.

Intervention design and resources

TAs reported that the prescription of the handbook, with detailed lesson plans and scripted questions, made delivering the intervention with fidelity straightforward, as did the provision of resources. As discussed in the Responsiveness section of this report, TAs commented that this had been supported by activities and games which made maths enjoyable for the pupils, and by the small-group nature of the intervention, which pupils benefitted from. Sufficient, regular time to deliver was seen to be key, as discussed above. TAs also liked the sequencing and pace of the intervention, which supported pupils to build on their learning as the intervention progressed and retained their engagement with the sessions.

Taking part in a trial

Link Teachers and TAs felt that delivering 1st Class@Number 1 through the trial had been a facilitator because it gave the intervention status and helped delivery to remain a priority. Trainers and school staff acknowledged that outside of the trial, time may not have been dedicated so consistently and other activities in school may have taken priority, which would have taken delivery off-track.

Support from trainers and EHU to support implementation in schools

The delivery team reported providing support, in addition to the virtual or in-person trainer visit, to TAs working in schools where the Link Teacher was unable to provide all the support needed. These TAs approached their trainer directly for support in relation to delivering the intervention, the maths content, and their interpretation of how to use the resources. The delivery team reported that this direct trainer support outside of sessions always forms part of the support package. During the trial, trainers also received queries from schools in relation to NFER's evaluation, which trainers sought guidance on from the delivery teams or from NFER to enable them to respond.

Perceived outcomes (RQ4.1, 4.2, 4.3, 4.4, 4.5)

RQ4: What were the perceived benefits of the intervention for TAs and pupils?

To establish the perceived benefits of the intervention, the research team asked survey respondents to comment on the extent to which they had recognised impacts the intervention aims to achieve, as detailed in the logic model, for TAs and for pupils. These impacts were also explored through the case study interviews and pupil focus groups in which Link Teachers, TAs, and pupils themselves fed back about how they felt they had benefited from participating in the intervention. Findings from across these data sources suggest that the intervention achieved the majority of the impacts outlined in the logic model.

Perceived outcomes for Link Teachers

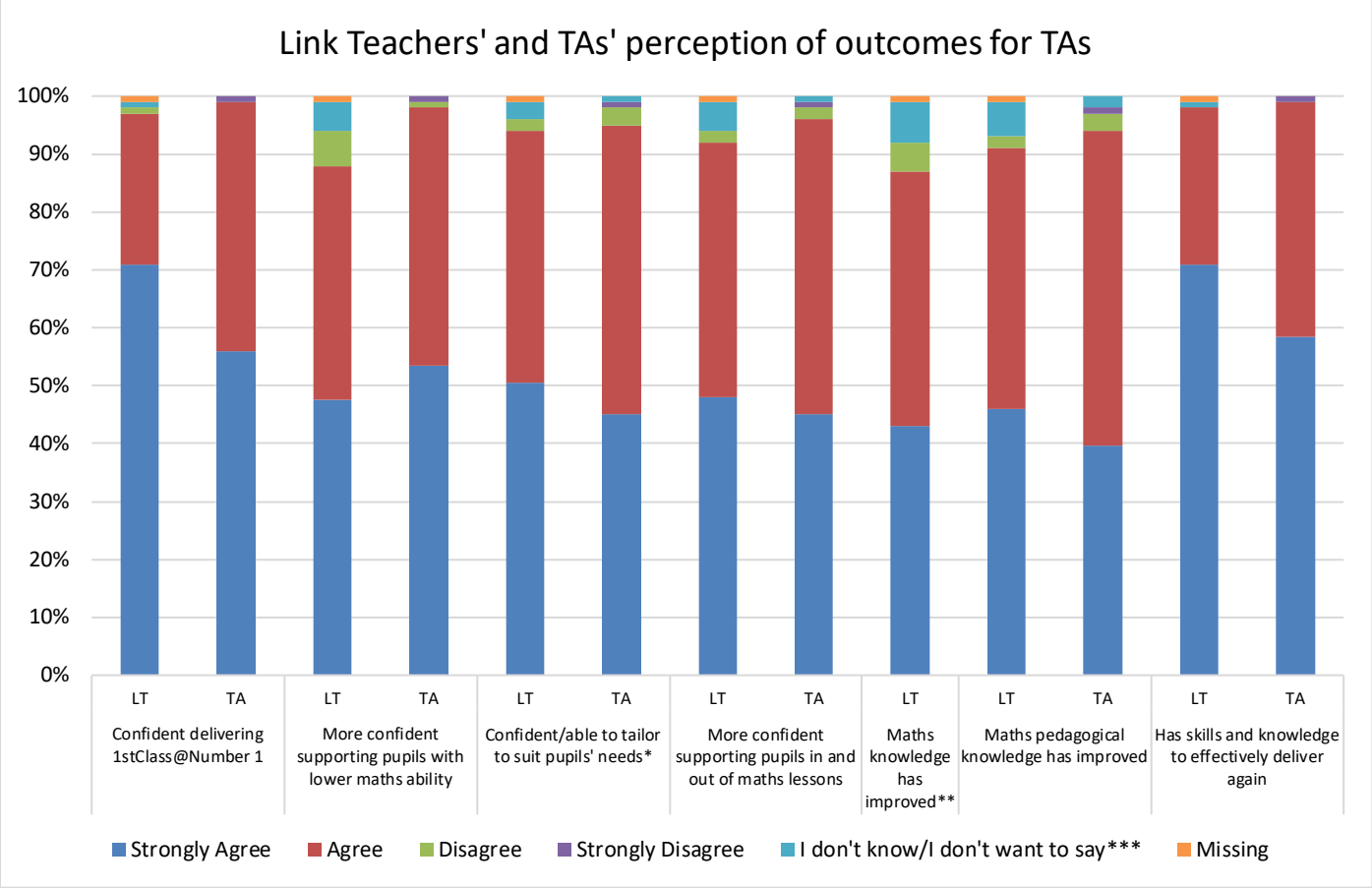
The 1st Class@Number 1 logic model states that the short-term outcome for Link Teachers is that they feel confident managing and quality-assuring the intervention. The research team asked case study Link Teachers to comment on any benefits they had seen for themselves resulting from their involvement. Due to their roles within middle and senior leadership, they already felt confident supporting the implementation of an intervention but appreciated having an in-depth understanding of the content and the progress that pupils were making. Link Teachers, who were maths leads but taught in Key Stage 2 also noted that they felt more aware of the Key Stage 1 maths curriculum and pedagogical approaches to securing the foundations of maths, which impact pupils as they progress through school. Link Teachers had also gained ideas from the final training session about how to support pupils not achieving ARE who had not secured basic maths skills and knowledge earlier on in school, which they felt they could implement in their own teaching, or share with other teachers, who taught these pupils. Link Teachers looked forward to sharing learnings from the intervention with other teachers throughout the school both informally and formally through CPD, for which materials were provided in the final training session. This anecdotal evidence provides some suggestion that the intermediate outcome for Link Teachers intended by the logic model—that Link Teachers understand how to implement interventions effectively, including the importance of liaising with staff within the school—had been achieved, at least within case study schools.

Perceived outcomes for TAs

The intervention logic model states that the short-term outcomes for TAs are improved maths subject and pedagogical knowledge, improved confidence teaching maths and delivering 1st Class@Number 1, and improved teaching practice. Intermediate outcomes relate to TAs ability to deliver the intervention independently and their confidence supporting pupils in and outside of maths lessons. The survey and qualitative data demonstrated positive impacts across these areas, suggesting the intended short term and intermediate outcomes of the logic model were achieved.

Within the survey, across knowledge-based statements, 87% to 98% of the 174 Link Teachers and 95% to 100% of the 187 TAs who responded to the surveys agreed or strongly agreed that the TA had developed in these areas. Across confidence-based statements, 87% to 97% of Link Teachers and 95% to 99% of TAs agreed or strongly agreed that the TA had improved in these areas. Figure 14 shows the extent to which Link Teachers and TAs agreed that each outcome had been achieved.

Figure 14: Link Teachers' and TAs' perceptions of outcomes for TAs



Data from the Link Teacher and TA survey: 'To what extent do agree with the following statements?' (Link Teacher N = 174, TA N = 187).
* Item worded differently across the two surveys. Link teacher survey: 'The TA is able to tailor the intervention to suit pupils' needs'; TA survey: 'I feel more confident tailoring maths activities to suit a range of needs/abilities.'
** Only Link Teachers were asked their views on whether TA's maths knowledge had improved.
*** Link Teachers were provided with an 'I don't know' response option; TAs with an 'I don't want to say' response option.

Data from the case study interviews adds further insights to the survey findings and aligns with the intended outcomes in the logic model. There was consensus among TAs that they had gained a better understanding of how pupils develop and learn core maths skills and were better able to teach and explain maths concepts to pupils both in the intervention and when working with these pupils in the classroom, both during and after intervention delivery.[45] TAs also reported a greater awareness and use of mathematical language when working with pupils more broadly because of the training they received during the trial. They felt more able to support lower ability pupils through identifying and correcting misconceptions as well as adapting their practice to suit pupils' needs, both during the course of the trial and afterwards. Link Teachers' views on how TAs had benefitted from the intervention aligned with TAs' own reflections on the impacts they had recognised for themselves. They also perceived TAs to have improved confidence in teaching maths and in leading an intervention. Link Teachers and TAs themselves reported that TAs had benefited from receiving their own CPD and the responsibility to lead an intervention had supported them to feel valued in school. There was also consensus across TAs and Link Teachers that the TAs would be confident delivering the intervention again and sharing their learning with colleagues. This is further discussed in Costs and Sustainability.

Perceived outcomes for pupils

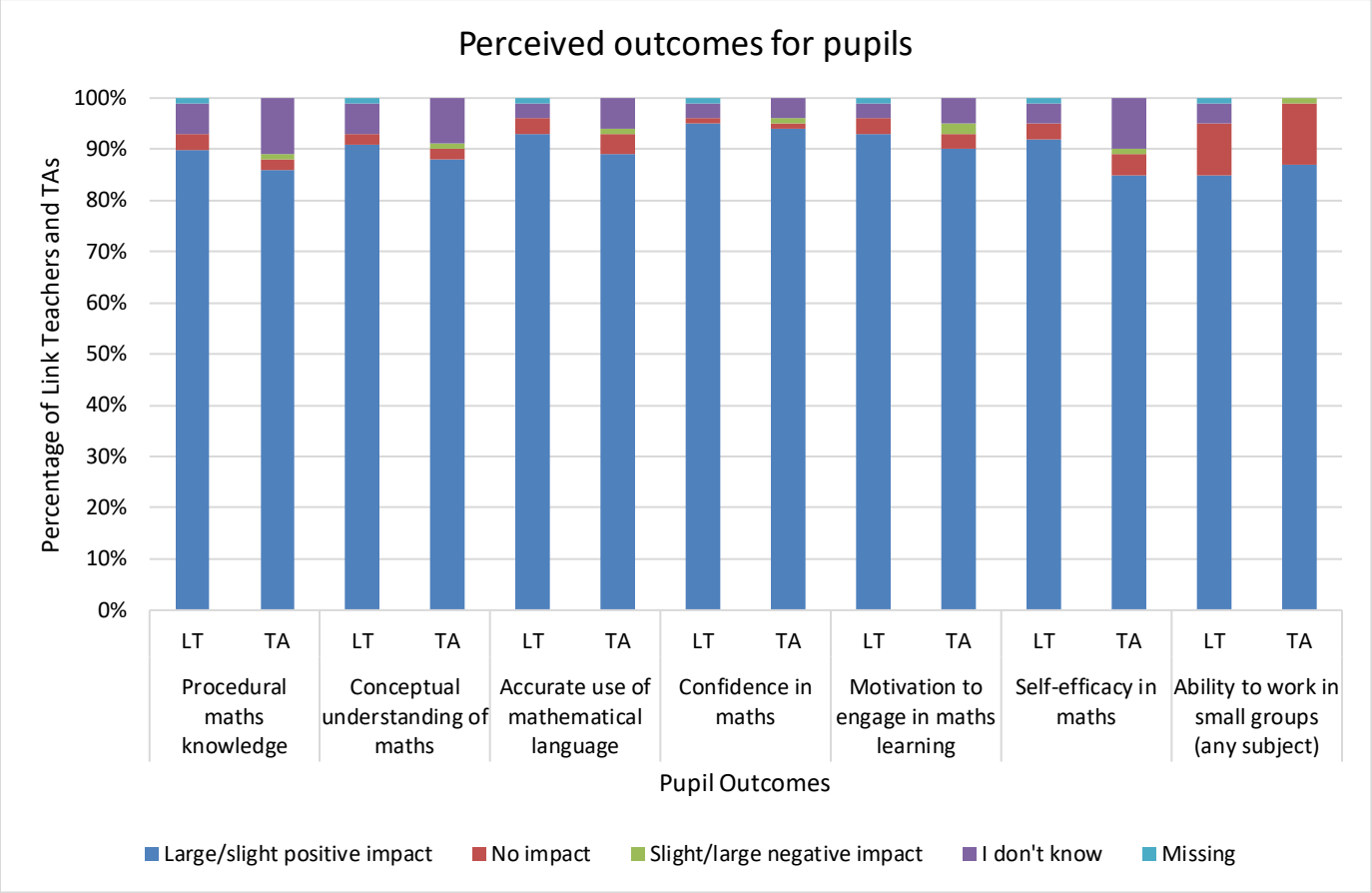
The 1stClass@Number 1 logic model states that the short-term outcomes for pupils are increased maths knowledge and understanding and improved use of mathematical language. Intended intermediate outcomes relate to pupils connecting

[45] Case study visits took place in the weeks following completion of intervention delivery, thus allowing TAs to comment on how receipt of the training and delivery of the intervention had informed their knowledge and practice in the classroom following delivery.

new and existing maths knowledge, demonstrating improved motivation and peer interactions, and experiencing success in maths. Findings from across the surveys and qualitative data activities are positive and suggest these intended outcomes were achieved for intervention pupils.

Survey responses indicated that nearly all of the 174 Link Teachers and 187 TAs who responded perceived that 1st Class@Number 1 had a positive impact on pupils' maths knowledge (86% to 92%) and on pupils' confidence in maths (85% to 96%). Fewer than 5% of Link Teachers and TAs reported no impact, or a negative impact, across each of the knowledge- and confidence-based items. The only exception to this was the item on pupils' ability to work in small groups where 10% of Link Teachers and 12% of TAs reported no impact or a negative impact. Figure 15 shows the extent to which Link Teachers and TAs perceived 1st Class@Number to positively impact pupils across a range of knowledge and confidence-based dimensions.

Figure 15: Perceived outcomes for pupils



Data from the Link Teacher and TA surveys: 'For pupils who have received the 1st Class@Number 1 intervention, what impact have you noticed on…?' (Link Teacher N = 174; TA N = 187).

Case study interviews provide further details on the perceived impact for pupils and also indicated that most of the intended outcomes as detailed in the logic model were achieved. Interviewees reported a range of improvements in relation to pupils' maths knowledge. TAs focused on specific topic areas or skills where pupils had shown improvements related to the topic areas and activities covered in the intervention, such as pupils' 'teen/ty' distinction, knowledge of addition, subtraction, times tables, and their number formation. They also reported that pupils were accurately, and more frequently, using mathematical language, particularly in later sessions. Meanwhile, Link Teachers tended to speak more broadly of the positive impact on pupils, such as their better understanding of the number system, their use of problem-solving strategies and manipulatives, and their ability to connect learning across maths concepts. Link Teachers who taught the intervention pupils in class reported that they had seen pupils transferring learning from the intervention into their lessons. Other teachers who taught the intervention pupils had provided the Link Teachers and TAs with similar feedback. The logic model intends that improvements in pupils' metacognitive skills in relation to maths will be a further intermediate outcome of the intervention. There was no explicit evidence from interviewees that this had been achieved

but the comments from Link Teachers in relation to pupils' problem-solving skills and connecting of maths concepts may suggest that the intervention could help achieve this outcome. Several Link Teachers also reflected on progress or assessment data when discussing the positive impact for pupils. They commented that intervention pupils had achieved ARE and felt that 1stClass@Number 1 had contributed to this.

> *'I just feel like it has been really successful. It has had such a strong impact on the children, which was far beyond what we thought it was going to; you could see it from the first few weeks, back in class things we were starting to see things click into place for them but also it highlighted where their gaps were. For example, there were a couple of them where we thought they were fine with numbers up to 20 so we had to go back but highlighting those gaps and providing a way to close them and give them the knowledge they needed was fantastic' (Link Teacher, case study school).*

The Link Teachers and TAs interviewed also commented on the positive impact 1stClass@Number 1 had on pupils' confidence. There was consensus across the interviewees that pupils were more willing to have a go and had an improved attitude towards maths. The small group nature of the intervention, which provided pupils with a supportive space in which they could confidently answer, even if they got a question wrong, was seen to facilitate these outcomes. Link Teachers and TAs also reported that improved confidence was transferring into the classroom: previously, these pupils would have remained quiet and let other pupils answer, but class teachers reported that pupils were contributing more to class discussions. In cases where the programme topic had preceded the teaching of the topic in class, TAs reported that intervention pupils felt like the 'experts' and were keen to share their understanding with their teachers and peers. This further contributed to their increased self-efficacy in maths.

> *'The biggest impact we have seen has been in the classroom, their contributions during whole-class teaching. They feel they are more confident to attempt things. Their attitude towards maths has improved. These were four pupils who felt they couldn't do it, their self-esteem in maths wasn't good and now the fact they are contributing in whole-class sessions is a marked improvement and getting the confidence is really important' (Link Teacher, case study school).*

In line with the intermediate outcomes related to improved peer interactions, Link Teachers and TAs also noted positive social impacts for pupils. Link Teachers and TAs reported that through the intervention, pupils had often worked with peers who they would not normally sit with in lessons or play with in the playground. This was particularly the case in multiple form entry schools where the intervention pupils came from different classes. Consequently, pupils had gained new friendships through the intervention and had been seen supporting each other back in the classroom or playing together at breaktimes. In line with the survey findings in relation to pupils' ability to work in small groups, TAs reported improvements in pupils' ability to take turns, listen to others, and work as a team.

Considering the data outlined above, there is considerable evidence that the 1stClass@Number 1 intervention was perceived to have positive impacts for pupils in relation to maths and school more broadly. However, the logic model also suggests that there may be short term and intermediate outcomes in relation to parent/carer involvement with the intervention and pupils' maths learning. As reported above in the Responsiveness section, parent/carer engagement was generally low, therefore it is unlikely that this contributed to intervention outcomes for most pupils.

Differences in outcomes

This effectiveness trial was commissioned to rigorously assess the impact of 1stClass@Number 1 with a particular focus on FSM-eligible pupils. Through the surveys and case study interviews, the IPE sought to understand any differential impacts for FSM and non-FSM pupils. However, it is important to note that the delivery team did not expect to see differences, therefore, FSM eligibility is not stated as an outcome moderator in the logic model.

Through the survey, around 60% of the 174 Link Teachers and around 40% of the 187 TAs who responded reported that all pupils had benefitted equally across each of the knowledge and confidence-based items. Around 25% of Link Teachers and 40% to 50% of TAs said they did not know if the intervention had a differential impact based on FSM, or that all intervention pupils were FSM eligible. Similarly, case study interviewees struggled to comment on FSM-related differences in impact. In some cases, they did not know which pupils received FSM or said that all intervention pupils were FSM-eligible. They reported that any differences in impact were based on pupils' starting ability, their appropriateness for the intervention, and their attendance. These factors align with the pupil-level moderators stated in the intervention logic model and suggest that intervention outcomes are not determined by FSM eligibility.

Unintended outcomes

The research team asked case study interviewees of any unintended outcomes they had identified for pupils. Class teachers had fed back that pupils were showing increased confidence in school more widely, for example, their confidence to answer in non-maths lessons and engage with their peers. Interviewees also reported that pupils had built a strong relationship with the TA and commented on the importance of pupils (particularly those from vulnerable backgrounds) having a trusted adult in school.

Pupils' own perceptions of outcomes

The pupil focus groups gave pupils in case study schools the opportunity to reflect on what they had learnt from their involvement in the programme by creating a learning tree to showcase their achievements. Pupils identified specific maths skills or content they had learned, which related to the 1st Class@Number 1 topics. This included counting skills, such as counting forwards and backwards from and to different numbers on the number line, addition, taking away, and recall of number bonds to five, ten, and 20. Pupils also reported that they could now count in twos, fives, and tens and could recall these times tables. Related to this, pupils were confident in their abilities to double and halve numbers. Pupils had gained familiarity with money and could confidently count coins of different values. Other skills included awareness and accurate use of symbols, 'teen and ty' numbers, and responding to questions using complete number sentences. Activities using a 100+ square had also given pupils an awareness of numbers exceeding 100. Figure 16 below shows a learning tree, developed by the research team during their analysis, to show the thematic groupings of the areas where pupils felt they had gained skills or knowledge.

Figure 16: Learning tree displaying topics which pupils felt they had learnt, or got better at, following 1st Class@Number 1

Context (RQ5.4, 7.1, 7.2, 7.3, 7.4)

RQ7: How manageable was it for schools to facilitate the pupil-randomised design?

School experiences of the trial

Overall, case study Link Teachers reported that they had found this trial easy to manage and would be willing to participate in future pupil-randomised trials. Some had already signed up for EEF trials running in the 2024/2025 academic year (although not exclusively pupil-randomised trials). A small number of case study school staff questioned the endpoint test selection (Quantitative Reasoning Test): they had concerns that, due to the test content not fully aligning with the topics and skills taught in 1stClass@Number 1, it would not fully capture the impact of the intervention compared to the SENT-R used to screen pupils. Some Link Teachers reported using the SENT-R assessment themselves to assess pupils' progress resulting from the intervention (this assessment is recommended by the delivery team for measuring pupil outcomes).

Adherence to trial

Adherence to the trial randomisation among schools was high. As reported in Compliance and Dosage, less than ten control pupils for whom we received maths intervention records were reported to have received 1stClass@Number 1. As noted above, these pupils came from three different schools and received the intervention due to an accidental switch of pupils by the schools in two cases and 'unforeseen circumstances' at the other school.

Staff interviewed through the school-based qualitative data activities were conscious of being part of an RCT and the importance of adhering to the randomisation. They reported avoiding contamination of the control group through delivering the intervention in a separate room or space outside of the Year 2 classroom.[46] They were also conscious of not sharing any resources with Year 2 teachers, or using materials from the intervention themselves, to support lower ability pupils during maths lessons. However, they were looking forward to doing this at the end of the trial. The research team asked TAs to comment on any changes they made to their practice resulting from the training, which may have influenced how they worked with control group pupils during the trial period itself. TAs reported feeling they were better able to model concepts to pupils using practical resources and were better able to recognise and support misconceptions. However, they had been careful not to refer to the intervention activities when using such strategies to support control group pupils in the classroom during the trial period.

Challenges with using a pupil-randomised design

The main challenge TAs encountered during the trial was managing pupil dynamics within the intervention group (see Facilitators and Barriers). These issues may have arisen due to the TA or Link Teacher not being involved in the screening or selection of pupils or being unaware of the guidance around switching out pupils they felt to be unsuitable for the intervention. Link Teachers and TAs acknowledged that, outside the trial, there were some pupils they would not have grouped together due to differing abilities and temperaments. However, their pairing strategies helped reduce challenges posed by the randomisation. For example, some paired pupils based on ability so that more able pupils worked together on tasks which were adapted to challenge them, which they could do with less support from the TA, while less able pupils worked together on tasks which were made easier, with additional TA support. Through the qualitative data activities, the research team identified that some TAs had delivered the intervention twice each day to pairs rather than the group of four. This was done where it was felt one to two support would be more beneficial than one to four—for example, where ability differences were stark, or where EAL pupils required more intensive support. This approach had been advised by trainers for these specific cases and if TAs had the time and flexibility to deliver in

[46] Although control group pupils in case study schools were not in the room when intervention group pupils received 1stClass@Number 1 and did not receive any of the intervention, there is the potential that they would have seen displays of resources and pupils' work if they received another intervention in the same room as it was observed these displays were left visible so other staff or pupils could potentially have seen them. This was contrary to the 'dos and don'ts' list provided to schools as part of the evaluation and is therefore a possible additional source of contamination, though we cannot determine the scale of this.

this format. There were also instances of pairing pupils based on behaviour and prior relationships (that is, not pairing pupils who may argue or distract one another).

Scale-up (RQ6.1, 6.2)

RQ6: What are the facilitators, barriers, and features of delivering at scale and is the intervention sustainable?

To understand how the intervention had been delivered at scale for the trial, the research team spoke with a member of the delivery team following the conclusion of the trial. This interview explored their experiences of delivering at scale, factors which had supported or hindered this, and their plans for delivering the programme at scale in the future.

Experiences of delivering 1st class@Number 1 at scale

It should be noted that EHU are already familiar with delivering 1stClass@Number 1 at scale. The delivery team reported that the number of schools they would usually deliver to in an academic year varies based on demand and schools' willingness and ability to pay for the programme. In a typical year, they could deliver to 150 schools—more if there was the demand. That said, the trial was somewhat larger than this as 226 schools were recruited to participate in the trial to ensure it was sufficiently powered to detect impact.

Delivery at scale is supported by having a network of trainers across the country. The delivery team reported that the trainers are experienced maths consultants who have worked with Every Child Counts for several years and must have delivered the 1stClass@Number 1 intervention themselves and been accredited by Edge Hill University before delivering the training. The steps taken by the delivery team to ensure scaled delivery is consistent and of high quality are discussed in Fidelity and Adaptations.

Initial challenges with delivering at scale through the trial were identified in relation to the recruitment of schools. The focus of this trial was on understanding the impact of 1stClass@Number 1 for FSM pupils. The funding for this trial came from the DfE's Accelerator Fund, therefore, EEF criteria stipulated that 50% of schools involved were in education investment areas (EIAs, areas of high levels of Pupil Premium needs). This meant that some of the areas EHU initially identified, based on where trainers were located, could not be involved in the trial. Consequently, they had to be flexible with the training groups and establish groups in other regions to meet the trial requirements. The online training group was created to accommodate schools that did not fall into a specific training region or were in a region where the minimum number of schools required to establish a regional training groups could not be recruited. In addition, the trial was delivered over two cohorts in order to deliver the intervention at scale to over 220 schools. The delivery team also reported some challenges in ensuring that training venues were located centrally across the group of TAs, acknowledging that TAs had to get back to school either for afternoon lessons or to pick up their own children from school.

EHU's plans for future training and delivery at scale (RQ6.2)

The delivery team reported plans to continue scaling-up their delivery of 1stClass@Number 1. They reported that their current network of trainers across regions will facilitate this, but to support further scaling and to ensure sustainability of the programme, they plan to continue to train and accredit suitable additional trainers. This increases capacity as more trainers enables the running of more training groups and more schools to be involved, but it also safeguards against current trainers moving on.

Delivery of online training will form part of EHU's future delivery model. However, the concerns raised about delivery of online training through the trial, such as trainers' ability to use the functionality of the Teams or Zoom platforms and the level of engagement experienced by delegates, will be addressed by the intervention's development leads to ensure online delivery is consistent, both in content and quality, to the standard achieved at in-person training. The delivery team reported that outside of the trial, development leads conduct visits to trainers' groups to monitor the consistency and quality of delivery. Following observations, development leads hold a professional learning conversation with the trainer which allows for reflection on the training and action points are discussed. Trainers who deliver online sessions in the future will receive the same form of support, which will allow the delivery team to closely monitor the sessions and ensure they are high quality.

The delivery team identified several factors which could support, or hinder, future delivery of the intervention at scale. They reported that under the previous government there had been a rhetoric of pupils 'keeping up not catching up', which had reduced schools' take up of interventions like 1stClass@Number 1. However, following Covid-19, the need for catch-up support was apparent and demand increased. The delivery team hoped that the new Labour government's guidance and policies around the use of interventions would return to supporting the use of such interventions in schools, which would enable them to continue delivering the programme at scale.

The delivery team highlighted that they felt the uniqueness of the 1stClass@Number 1 programme, compared to programmes developed by competitors, was its coverage of content aligned with the Year 2 maths curriculum and the provision of materials and resources. While this makes the intervention attractive to schools, the delivery team acknowledged the financial challenges schools are facing and expected schools' ability to fund the intervention to be a barrier to them delivering the programme to more schools. Additionally, they commented on the staffing challenges that schools are facing. In the current context, many schools are having to reduce their TA numbers to save money, which in turn reduces their capacity to offer small group interventions. In schools where TAs are retained, the delivery team commented on increasing staff sickness which impacts TAs' ability to attend the training and then deliver the intervention as they are required to cover for absent colleagues.

Costs and sustainability (in schools) (RQ6.3, 6.4, 6.5)

Data from the surveys and case study interviews were used to ascertain the extent to which continued delivery of the intervention was perceived to be sustainable within schools, with consideration given to the costs, resources, and time required to support this. The research team also explored what delivery may look like in schools in the future as well as facilitators and barriers to continued implementation.

Sustainability of 1st class@Number 1 in schools

Data from surveys and interviews are positive in relation to the sustainability of the intervention in schools. Most of the 174 Link Teachers (89%) who responded to the survey reported that they planned to continue delivering 1stClass@Number 1. No Link Teachers reported in the survey they would not be continuing with the intervention, but 10% were unsure.[47] Link Teachers were asked to indicate their top three reasons for continuing to deliver the intervention. The percentage of teachers who selected each reason are detailed in Table 23.

Table 23: Reason for continuing to deliver 1stClass@Number 1

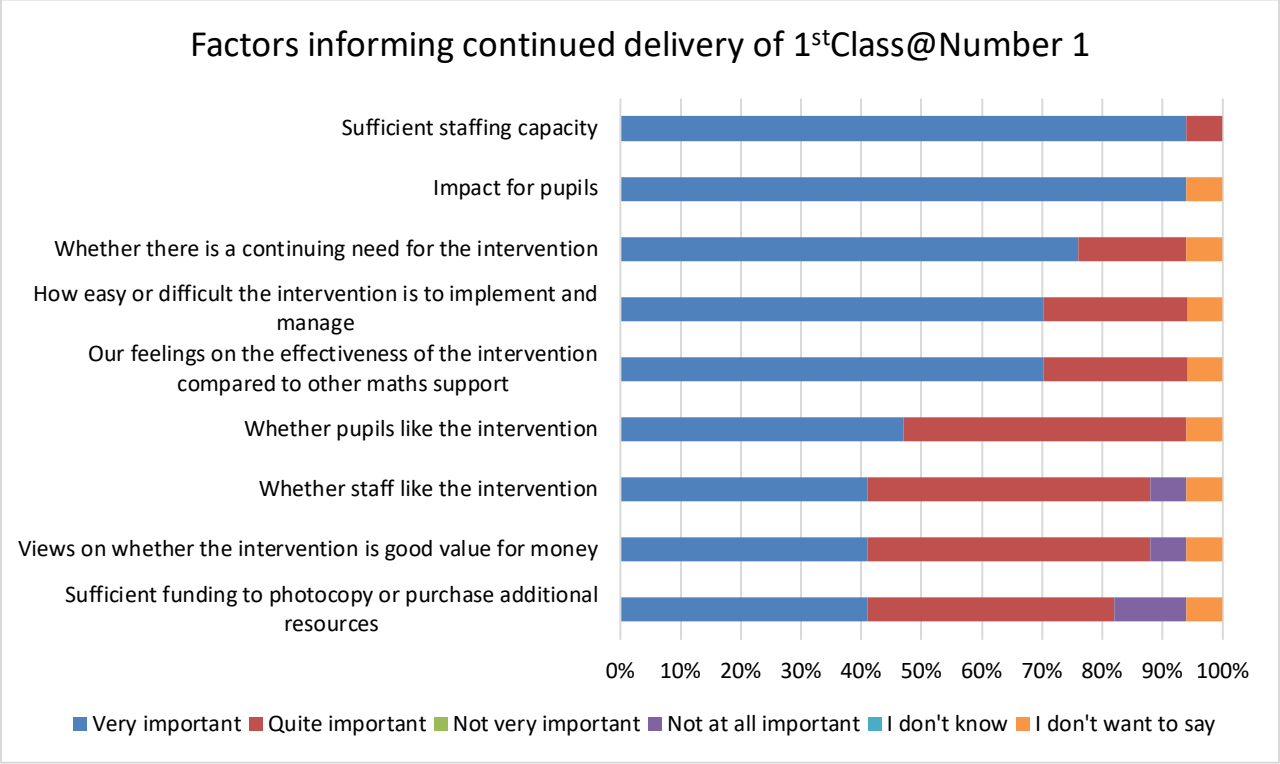
Reason for continuing to deliver 1stClass@Number 1	Percentage of Link Teachers (%)
There continues to be a need for the intervention	72
We are pleased with the impact for pupils	68
The intervention is easy for us to implement and manage	38
Our pupils like the intervention	38
We feel this intervention is more effective than other maths support for improving pupils' outcomes	29
We have the staffing capacity to continue delivering the intervention	19
Our staff like the intervention	16
We think the intervention is good value for money *	10

* In this trial, schools received the programme at a discounted rate—see Value for Money section.

[47] The remaining 1% of responses were missing.

Similarly, most Link Teachers in case study schools planned to run the intervention again. Aligned with the survey findings, the main reasons they reported for continuing delivery were the perceived positive impact for pupils, the continuing need for the intervention, and it being a programme which pupils enjoyed participating in. In general, Link Teachers felt the time and resources required to deliver the intervention in the future was sustainable and worthwhile. The majority of TAs who responded to the survey indicated that, if their school continued to deliver 1st Class@Number 1, they would like to continue delivering the intervention (89%). No TA reported in the survey that they would not like to deliver the intervention again, but 10% were unsure.[48] Those who were unsure were asked to comment on the importance of a range of factors for informing their decision. Across each of the factors presented in the survey, at least 80% of these Link Teachers said these would be 'very' or 'quite' important factors, the most important appearing to be whether schools had sufficient staffing capacity and the impact identified for pupils (Figure 17).

Figure 17: Importance of factors for informing continued delivery of 1st Class@Number 1



Data from the Link Teacher survey: 'How important will the following factors be in informing your decision?' (N = 17). Items from the survey question have been reordered to be presented in descending order.

Among case study Link Teachers who were uncertain over future delivery of the intervention, the balance of the time it takes to deliver the intervention against the impact seen for pupils was the key factor for informing whether delivery would continue. They reported that having a TA out of the classroom for 40 minutes or more was a significant amount of time and they felt the scale of the impact (that is, for four pupils) was relatively small (note, they are not commenting here on the actual impact results observed in the trial). Link Teachers reported that if they chose not to deliver the intervention in the future, they would revert to delivering the interventions reported under Usual Practice.

Future delivery of 1st class@Number 1 in schools

Of the Link Teachers who were certain they would continue delivering the intervention (N = 154, 89%), just over three-quarters (77%) reported they would deliver to Year 2 pupils in the following academic year. Just less than half said they would deliver to pupils who they screened but were not included in the trial (47%) and just over two-fifths said they would deliver the intervention to control group pupils (42%). This latter proportion appears low, particularly given one of the

[48] Note the remaining 1% of responses were missing

benefits of the pupil-randomised trial is that eventually even pupils assigned to the control group can receive the intervention. However, as noted below, case study interviewees in Cohort 2 did not have sufficient time to deliver the full intervention before the end of Year 2 and interviewees across cohorts acknowledged that there were other pupils who would benefit from receiving the intervention, not just control group pupils. A smaller proportion (16%) said they would offer the intervention to other pupils, including Year 2 pupils who had not initially been screened or put forward for the trial, but whom teachers perceived to need additional support. Several Link Teachers reported that they planned to deliver the intervention to low attaining pupils in Year 3 upwards who had not yet met Key Stage 1 objectives. Link Teachers added that they would use progress and assessment data to inform their pupil selection.

Findings from the case study interviews showed that the timing of the original intervention influenced schools' future delivery plans. Cohort 1 case study schools had time to start the intervention again, after their endpoint testing, with another group of Year 2 pupils in the 2023/2024 academic year. They reported plans to deliver to control group pupils or other pupils identified to still be working below ARE. Cohort 2 case study schools spoke of their plans to deliver the intervention in 2024/2025 to groups within Year 2, to control group pupils or other pupils identified as working below ARE once they moved into Year 3. They acknowledged that although the intervention is targeted at Year 2 pupils, they had identified a group of pupils who would still benefit from receiving 1st Class@Number 1 upon moving into Year 3. Schools also had plans to deliver the intervention to groups of six pupils or hoped to run multiple groups, either concurrently or throughout the academic year, to maximise the impact that could be achieved. Link Teachers said that they would select and group pupils based on ability, pupil relationships, and suitability for the intervention.

All case study schools that planned to deliver the intervention again had plans to cascade the training to other TAs in school (which would facilitate the running of multiple intervention groups) and to Year 2 teachers, who could implement strategies and activities in their maths teaching where they felt appropriate. This was also seen to safeguard against the trained TA or Link Teacher leaving the school and taking the knowledge of how to deliver the intervention with them. We note, however, that the training cascading model has not been tested and would be a significant shift from the training model that has been tested in this effectiveness trial. Training delivered by TAs who have not been trained by EHU's accredited trainers may not be as effective in improving pupil outcomes.

Several schools reported that in the future they would use the SENT-R to conduct their own baseline and endpoint assessments to help them measure the impact of the intervention. The use of the SENT-R at baseline and endpoint has also been previously recommended by the delivery team.

Facilitators and barriers to continued delivery in schools

As reported in Perceived Outcomes, almost all of the 174 Link Teachers (98%) who responded to the survey reported that they felt the TA had the skills and knowledge to effectively deliver the 1st Class@Number 1 intervention in the future and all TAs agreed with this statement. When asked what could facilitate continued delivery in case study schools, Link Teachers identified the TAs' skills and knowledge as a key facilitator. Furthermore, TAs themselves reported that they would be confident delivering the intervention again. However, staffing capacity was identified as a key factor which may facilitate, or hinder, future delivery. Case study staff expected that the TAs' time for the intervention could remain on their timetables but highlighted that issues could arise if school priorities changed and the TA was required to deliver other interventions, or if they could no-longer be released from their classroom responsibilities.

Link Teachers felt that the provision of resources would also facilitate future delivery. They reported that they now have master copies of all materials, which can be photocopied, and they appreciated that they received access to a portal to download additional copies. They did, however, comment that they would find it beneficial to have digital copies of the topic handbooks too as, over time, these would become worn.

Link Teachers reported that school budgets would be a barrier to employing more TAs to deliver the intervention, or to allowing other TAs to attend the 1st Class@Number 1 training, hence their plans to upskill other TAs internally.

Value for money

Through the survey, almost all of the 174 Link Teachers (92%) who responded reported that they believed the 1st Class@Number 1 intervention offered good value for money. However, many reported that had they not received the intervention at a subsidised rate (£200) as part of the trial, they would not have been able (67%) or not willing (6%) to pay £1,100 for the intervention. Just 6% of Link Teachers reported that they would have been able to fund the full cost

of the intervention. Interviews with Link Teachers in case study schools sheds further light on these perceptions. Views over value for money were mixed. Link Teachers in case study schools who were senior leaders, or familiar with making financial decisions regarding interventions, felt better able to comment on this. Most felt that 1stClass@Number 1 did offer good value for money given the comprehensive training, resources, and the impact they had recognised. However, they also acknowledged that the cost of the intervention versus the scale of impact (that is, one TA, four pupils) did reduce the value for money. In line with the survey findings, there was consensus from case study Link Teachers that their schools would have been unable, or would have struggled, to fund £1,100 for the intervention outside of the trial.

Cost

A full cost evaluation was performed by Nunes et al. (2018) in the previous evaluation and the delivery of 1stClass@Number 1 has remained mostly unchanged since then. We therefore followed this same costing framework as far as possible, updating costs using figures obtained for 2023/2024 (where available) or by applying inflation adjustments. The costing assumptions are summarised below, including any changes since the previous evaluation.

Number of pupils

In the previous evaluation, 83% of schools planned to deliver or were delivering 1stClass@Number 1 to a second group of pupils in the same academic year. On this basis it was assumed that eight pupils would receive the intervention each year (two groups of four pupils) for the purposes of costing, for a total of 24 pupils over a three-year period. We also use eight pupils per school per year for costing, based on these previous assumptions, rather than any data about the current trial pupils. In theory it is possible to deliver 1stClass@Number 1 to 12 pupils a year (one group per term), so we also included a cost calculation using this assumption as a sensitivity analysis. However, eight pupils per year was used to estimate cost in the previous trial, so this is the number we used for our main cost estimate in Table 24.

Training

The cost of training one TA and one Link Teacher in 1stClass@Number 1 is £1,100 per school as of 2023/2024, which includes a pack of teaching resources and printed resources for four children. For the main costing (Table 24 below) there were assumed to be no costs to the school for transport to the training or supply cover during it, as in the previous evaluation. However, the cost of full supply cover for the LT to attend training was included as a sensitivity analysis.

Additional resources

In the 2018 evaluation, the cost of additional resources needed to deliver the intervention (such as counters or cubes) was estimated at £31 for the first group of four pupils in a school and then £15.50 for each subsequent group, based on the average costs reported by schools. These costs have been adjusted in line with inflation between July 2018 and July 2024 (measured using the Consumer Price Index), which amounts to a 26.4% increase. The updated costs are £39 for the first group of pupils and £19.50 for subsequent groups.

Optional costs

The charge suggested by EHU for a trainer to visit a school, to observe and provide one to one support for the TA, is £270 as of 2023/2024. This visit is optional in business as usual, but was considered valuable by schools so all schools received it as part of the trial and it is included in the total costs in Table 24.

EHU provide an ‘online resources and data collection’ service: schools can track the progress of pupils, as well as access further information and downloadable resources via the EHU ECC website. This is free while receiving the 1stClass@Number training but costs £162 per year in subsequent years (as of 2023/2024). This cost is optional so it is listed in Table 24 but not included in the final total.

For the previous evaluation, TAs could take an online distance study module at EHU to gain undergraduate university credit, which was included as an optional cost. This option is no longer offered by EHU so is not included anywhere in our costs.

SENT-R

Costs for the SENT-R were obtained from GL Assessment, using costs for 2023/2024. We assumed that most schools would want to use the SENT-R both for screening to select the most appropriate eight pupils for 1st Class@Number 1 and as a pre/post test to monitor progress. This would require 20 SENT-R Form A booklets (for screening and as a pre-test)[49] and ten Form B booklets (for the post-test) each year. Booklets are purchased in packs of ten, with each pack costing £36.60 (including VAT).

In the first year, the cost of using the SENT-R package is £302.40 (including VAT), which includes ten Form A and ten Form B booklets as well as access to the online marking tool. A further ten Form A booklets are also required, so the total cost of the SENT-R in the first year is £339.

In the second and third years, schools need to purchase 30 extra booklets for £109.80 and the online marking tool separately at a cost of £85.20 (including VAT), which totals £195 per year.

Table 24: Costs of the 1st Class@Number 1 programme for two groups of pupils each year over a three-year period

Cost	Year 1	Year 2	Year 3
1. Training, teaching materials, and printed resources	£1,100	£0	£0
2. Additional resources per group (ongoing)	£58.50 (£39 + £19.50)	£39 (2 × £19.50)	£39 (2 × £19.50)
3. One to one support (one-off)	£270	£0	£0
4. Optional online resources and data collection (ongoing)*	£0	£162	£162
5. SENT-R	£339	£195	£195
Total (1+2+3+5)	£1,768	£234	£234
Cumulative total	£1,768	£2,002	£2,236

Figures may not add up exactly due to rounding.
*Cost (4) is considered optional and so is not included in the cost total

Assuming 8/16/24 pupils have received 1st Class@Number 1 by the end of the first/second/third year, the average cost per pupil drops from £221 in the first year to £125 by the end of the second year and then to £93 by the end of the third. As a sensitivity check we also considered a case where schools delivered 1st Class@Number 1 to 12 pupils each year rather than eight. In theory, it is possible for schools to complete delivery to the first group of four pupils before Christmas, leaving time for two additional groups during the remainder of the academic year. This would add about £19.50 per year for cost (2) in Table 24 and £36.60 in the first year for cost (5).[50] The cumulative total therefore rises compared to Table 24, to £1,824/£2,078/£2,331 in the first/second/third year, but the average cost per pupil falls to £152/£87/£65.

In the previous evaluation both TA and lead teacher time to attend training was assumed to be covered internally and so was not represented as a financial cost. As another sensitivity check we also considered a scenario where lead teacher time to attend training was covered externally by a supply teacher. This would require approximately one day

[49] We assume schools will screen somewhere between 11 and 20 pupils. More than 20 pupils would require extra Form A booklets to be purchased.

[50] We assume that schools will have to buy an additional ten Form B SENT-R booklets in the first year due to testing 12 pupils instead of eight as they are sold in batches of ten. In theory this should leave eight spare booklets, which is the number needed to cover the extra four pupils tested in the second and third year. Form A purchasing is unaffected: it is assumed no more than 20 booklets will be needed for screening.

of cover (for two half-day training sessions) incurring a one-off cost of £218 on average.[51] The total cost at the end of the first year then increases to £1,986 and the cumulative cost over three years to £2,454. We considered it unlikely that TA training time or lead teacher time supervising the TA would be covered externally, so have not included this time as a financial cost.

[51] Based on Use of supply teachers in schools: research reports, Table 3.

Conclusion

Table 25: Key conclusions

Key conclusions	
1.	Pupils receiving 1stClass@Number 1 made the equivalent of two additional months' progress, on average, compared to pupils who did not receive it. This result has a high security rating.
2.	Pupils eligible for free school meals (FSM) receiving 1stClass@Number 1 made the equivalent of two additional months' progress, on average, compared to FSM-eligible pupils who did not receive it.
3.	TAs perceived the training content, delivery, and materials to be of high quality and reported that they enjoyed delivering the intervention to pupils. In over 80% of schools, the same TA attended all of the delivery-training sessions, that is, there was continuity of TA, which is important for effective delivery.
4.	Pupils reported enjoying the intervention activities and interactive games. TAs and Link Teachers felt pupils were engaged and that the programme had a positive impact on pupils' maths attainment. Just over 60% of pupils attended at least five of the six sessions for each topic—a threshold which was felt to be important for the success of the programme.
5.	The key features of 1stClass@Number 1 that are likely to have led to the positive impact on pupils' attainment include its modular nature with each topic building on previous ones, its manualised nature, high quality training, high quality small group tutoring with optimum group size and session frequency, and the ability to be adapted to pupils' needs.

Impact evaluation and IPE integration

Evidence to support the logic model

The logic model for 1stClass@Number 1 includes each group of participants (TAs, Link Teachers, pupils, and parents/carers) as separate rows allowing us to represent the moderators and the short-, intermediate-, and long-term outcomes for each group.

TA selection is an important moderator in the intervention logic model and is represented in terms of TAs' qualifications and experience delivering maths interventions and, more broadly, any interventions in KS1. However, the results from the regression models exploring the moderating effect of each of these variables provided no evidence that they moderate the impact of 1stClass@Number 1 on pupils' maths attainment measured via the QRT. That said, most Link Teachers were satisfied that they had selected the most appropriate TA to deliver 1stClass@Number 1, which may indicate that there were factors we did not, or could not, measure that are important for effective delivery or that interact with TA knowledge and experience. Alternatively, it is also possible that the selection criteria and process for TAs meant that there was not enough variation in knowledge and experience to establish a relationship.

We also explored other potential moderators relating to TAs, Link Teachers, pupils, and schools through the IPE. These included delivery and preparation time for TAs, support provided by Link Teachers, pupils' needs and group dynamics, support from other school staff, alignment of 1stClass@Number 1 with maths curriculum, and implementation support provided by EHU to schools. Preparation time for TAs was perceived as an important moderating factor: when this was secured in timetables (either in the form of more than three weekly 1stClass@Number 1 sessions or by allocating a longer duration of time for the delivery of each session), TAs were able to spend the additional time flexibly by either preparing for the sessions, revisiting content with pupils requiring additional support, or helping pupils catch-up with missed sessions. Alignment of the 1stClass@Number 1 sessions with the maths curriculum may also be an outcome moderator: Link Teachers and TAs reported that the content of the 1stClass@Number 1 sessions often aligned with the topics covered in class, which meant pupils could draw on their learning from the intervention to support them in maths lessons or vice-versa. However, these comments were more frequently made by Cohort 1 staff and, as demonstrated by the survey findings, there were instances of Cohort 2 staff reporting that the topics would have aligned better had they started the intervention earlier in the academic year due to the timing of topic delivery in the Year 2 maths curriculum.

A key assumption of the intervention logic model was that appropriate pupils would be selected to receive the intervention. Pupil selection was also highlighted as an area of focus in BIT's formative evaluation following the previous

effectiveness trial: that evaluation concluded that schools might have selected a 'not unsubstantial' number of pupils who did not need the intervention, which could have impacted the outcomes observed. In response to this finding, we developed and followed an objective screening process in this trial using the SENT-R to identify pupils suitable for the trial. Of the 1,797 pupils in the trial, only 15 (< 0.01%) had SENT-R scores above 49, indicating that they were not experiencing moderate difficulties with maths and therefore did not need 1stClass@Number 1. This suggests that our approach to pupil selection for the trial mitigated, to a large extent, the issue of pupils receiving the intervention when they did not need it. The selection of these 15 pupils was likely a consequence of the pupil-randomised design wherein schools were asked to nominate at least eight pupils to the trial. In smaller schools this might have meant that schools nominated pupils who were not suitable for the intervention in order to maintain fidelity to the design of the trial. While the selection of pupils who do not need the intervention does not appear to be a significant issue, the trial may have also selected pupils who were experiencing severe difficulties with maths, making them unsuitable candidates for 1stClass@Number. This was a consequence of not setting a lower bound on the SENT-R scores.

There is considerable evidence from the IPE to suggest that the short- and intermediate-term outcomes for TAs, Link Teachers, and pupils had been achieved. For example, TAs reported improved knowledge and confidence teaching mathematics as well as a greater awareness of correct mathematical language. TAs, Link Teachers, and pupils reported that pupils' maths knowledge, skills, and confidence (both in the intervention and in the classroom) had increased. TAs and Link Teachers also reported increased (and accurate) use of mathematical language by pupils towards the later sessions and the ability of intervention pupils to connect learning across maths concepts and transfer this learning from the intervention into their maths lessons.

Parent/carer engagement with the intervention appeared low with school staff indicating that only around a third had engaged with the special delivery activities. Therefore, while it is likely that parent/carer engagement might have helped some pupils, more systematic observation or measurement of the prevalence of this engagement might have helped us understand its impact on pupil engagement and attainment.

Because of this trial's focus on FSM pupils, TAs and Link Teachers were asked to comment on any differential impact they observed for FSM pupils receiving the intervention when compared to similar non-FSM pupils. Consistent with the delivery team's expectation, there was no clear evidence to suggest that the intervention resulted in differential impacts for FSM pupils. This is also in line with our subgroup analysis, which revealed similar effect sizes for FSM and non-FSM pupils receiving 1stClass@Number 1 in comparison with control pupils in each subgroup. While the previous trial found no evidence of impact on FSM pupils, in this trial, which was well-powered for FSM, we found that the intervention was essentially equally as impactful for FSM pupils as their non-FSM peers. Rather than FSM-eligibility status, TAs and Link Teachers reported that any differences in impact they observed were based on:

- pupils' starting ability—our subgroup analysis showed that the intervention was as impactful for pupils with low prior attainment, that is, SENT-R scores of 40 or less, as it was for all pupils;
- their appropriateness for the intervention; and
- their attendance at the intervention sessions.

These factors align with the pupil-level moderators stated in the intervention logic model.

The impact analysis supports the role of pupils' attendance at the intervention sessions and TAs' completion of the training sessions as implicit mediators of the intervention's impact although we recommend that the results of the compliance analysis be interpreted with extreme caution. The effect size for compliers among all pupils was quite high, ranging from 0.24 (for measure 2) to 0.36 (for measure 1).[52] Due to concerns around the violation of the exclusion restriction for compliance measures 1 to 3, the CACE estimate for these measures are likely to be upwardly biased. We determined the CACE estimate for a further compliance measure (the minimum compliance measure of pupils attending at least four of the six sessions in at least four of the five topics), that is, the highest threshold at which the exclusion restriction holds. Since 95% of pupils met this threshold, the CACE estimate for the minimum compliance measure (0.15) was similar to the ITT estimate for all pupils (0.12). Similar results were observed for the CACE estimates for

[52] The effect sizes for compliers range from two months' additional progress (minimum compliance, lower bound) to five months for pupils who complied with measure 1.

FSM pupils (0.11). The CACE estimate for the minimum compliance measure for FSM pupils was also comparable (0.14) to that of all pupils.

Interpretation

This is the second effectiveness trial of the 1stClass@Number 1 intervention. The first trial found an effect size of 0.18 (that is, an additional two months' progress) for all pupils. This result was not statistically significant likely because the trial was not adequately powered. In the previous effectiveness trial, the QRT was used both to select pupils into the trial (in combination with other eligibility criteria) and as a baseline measure. Consequently, the trial's baseline measure had its distribution curtailed as only children scoring below a certain threshold were selected. This correspondingly resulted in a much lower correlation between baseline and outcome (0.29) than was expected (> 0.7). To improve the QRT pre/post correlation in this trial, we used the SENT-R to select pupils to the trial and the QRT as the baseline and outcome measure. As expected, this strategy helped to increase the pre/post correlation of QRT from 0.29 to 0.4.

Pupils receiving 1stClass@Number 1 made the equivalent of two additional months' progress, on average, compared to those who did not. This is the best estimate of the intervention's impact, but a range of small and moderate positive effect sizes are also supported by the data to a lesser extent. Zero impact is not supported by the data, so it is extremely likely that the intervention improves maths attainment on the QRT. In contrast to the previous trial's findings, this trial was well powered for FSM-eligible pupils; FSM pupils receiving 1stClass@Number 1 also made the equivalent of two additional months' progress compared to control pupils.

Because this trial was run in two cohorts, we conducted a subgroup analysis to examine the differential impact of 1stClass@Number 1 on Cohort 1 and Cohort 2 pupils. The effect size for Cohort 1 was 0.14 and for Cohort 2 was lower at 0.08, although the finding for Cohort 2 was not statistically significant, likely due to the smaller sample in this cohort (70 schools in Cohort 2; 156 in Cohort 1). There are several possible explanations for this differential impact. Attendance at all five training sessions was lower for Cohort 2 TAs (56% for Cohort 2; 66% for Cohort 1) and fewer TAs from Cohort 2 attended each training session in comparison to Cohort 1. The reduced compliance for Cohort 2 TAs with training attendance is likely to have had an impact on the effect size as these TAs might not have been as well prepared to deliver the topics for which they missed training. Furthermore, whereas only five of the 156 Cohort 1 schools (3%) withdrew from intervention delivery, six of the 70 Cohort 2 schools (9%) did not deliver the intervention, thus likely reducing the effect size seen in this cohort. Another possible explanation is that Cohort 2 included one online training group for Link Teachers and TAs whose efficacy has not yet been tested: consistent IPE evidence suggests that it might not have provided the same high quality experience as the in-person training, which may have contributed to the lower effect size for Cohort 2. However, a sensitivity check restricting the primary analysis model to pupils whose TA attended 1stClass@Number 1 training at least partly in person, excluding those whose TAs were in the online training group, suggests this might not be the case. The effect size in this subgroup was 0.11, close to the primary analysis effect size of 0.12. The primary analysis result was therefore mostly unaffected by the inclusion of pupils whose TA attended purely online training, perhaps unsurprising given that there were only 50 such pupils. We also note findings from the IPE where TAs and Link Teachers reflected on the lack of alignment for intervention topics with the topics in the Year 2 national curriculum, particularly for Cohort 2 schools. This factor may have also contributed to the smaller effect size for Cohort 2 schools.

Given that pupils received 1stClass@Number 1 in addition to their usual maths lessons, and as part of small group tutoring, it is reasonable to ask whether the effects we have observed are related to pupils experiencing more maths and/or to high-quality small group tuition. We have attempted to determine whether the impact observed in this trial can be ascribed to 1stClass@Number 1 by examining the impact results in the context of the IPE findings and existing evidence on small group tuition in maths and specifically those delivered by TAs.

We examined the additional maths support offered to control pupils to explore whether (and how many) control pupils were offered small group support (that is, did we have an active control group?) and what this revealed in terms of the direct impact of 1stClass@Number 1 on intervention pupils. As described in the Usual Practice section, only 124 of the 658 control pupils (19%) received other additional maths support including formal interventions and small group support on specific topics. This data suggests that the majority of control pupils did not receive any additional maths support during the trial and did not constitute an active control group who spent additional time doing maths. This leaves open the possibility that the impact on attainment observed for intervention pupils could be a consequence of intervention pupils spending additional time receiving maths instruction.

The EEF Teaching and Learning Toolkit suggests that there is moderate evidence that small group tuition has a moderate impact on pupil attainment for low cost (EEF, 2021).[53] The evidence suggests that the impact of small group maths tuition is an additional three months' progress, on average, during the course of a year for primary pupils. There are several active ingredients that contribute to its effectiveness: (1) group size—the smaller the group, the closer the interaction and feedback between educators and pupils and, therefore, better impact on pupils' attainment; (2) frequency and duration of sessions—frequent sessions, three times a week for up to an hour delivered over ten weeks have the greatest impact; and (3) targeting support to pupil needs—the identification of pupils who are at risk of falling behind and targeting support based on each pupil's understanding can support pupils to overcome barriers to learning. The 1stClass@Number 1 intervention incorporates all these active ingredients of small group tuition that have been demonstrated to have a positive impact on pupils' maths attainment. Other differentiating features of 1stClass@Number 1 that might have contributed to this positive impact include the highly structured nature of the intervention, where each topic builds on previous ones; it is also an example of a highly manualised intervention where TAs are provided with extensive resources to deliver each of the six intervention sessions for each topic in a structured manner. While the intervention itself is very structured, the resources and training also provide TAs with strategies to adapt the intervention to suit the needs of pupils in the group thereby allowing TAs to be responsive to pupils' abilities and requirements. An additional mechanism through which 1stClass@Number 1 might have specifically impacted pupils' attainment is increased pupil engagement with the intervention. Our IPE revealed that TAs, Link Teachers, and pupils themselves felt that pupils were engaged with the content of the intervention—specifically the games and activities—which is unique to 1stClass@Number 1.

In addition to the above, TAs have been shown to have a positive impact on pupils' attainment when they deliver structured interventions with high quality training and support in one to one or small group settings (Sharples, Webster and Blatchford, 2021). The EEF recommends that when interventions are delivered by TAs, schools should use interventions that provide structured resources and lesson plans with clear objectives and ensure that fidelity to the intervention is maintained during delivery. The 1stClass@Number 1 sessions are highly manualised, providing TAs with comprehensive lesson plans, scripted questions, and resources. Link Teachers and TAs were satisfied that the content of the training sessions, resources, and support provided were of consistently high quality and prepared them well to deliver the intervention sessions. Furthermore, as detailed in the IPE findings, the intervention was largely delivered by TAs with fidelity, supporting the suggestion that the impact observed in this trial is a consequence of intervention pupils receiving 1stClass@Number 1.

The use of TAs to deliver a highly manualised intervention like 1stClass@Number 1 is an efficient and potentially cost-effective use of limited resources in schools. By providing high quality training supported by comprehensive resources, 1stClass@Number 1 allows TAs to independently deliver small group tutoring thereby allowing teachers to focus on classroom teaching. The intervention also builds capacity within schools for TAs to support pupils facing moderate difficulties thereby removing the need for potentially more expensive provision by external tutors.

Given this was an effectiveness trial, an exploration of sustainability and further scale-up was important. We found that most schools intended to continue to use 1stClass@Number 1, based on their positive experience of it and perceptions that it had improved the pupils' maths attainment. They felt there was a continuing need for supporting pupils' maths learning, and they intended to use it both for control group pupils where still possible within the academic year, and for Year 2 pupils in the following academic year. However, in order to make it more manageable in terms of resourcing (one TA to four pupils), many schools were intending to increase the group size to, say, six or more pupils. We recommend that EHU monitor such continued practice and seek feedback from schools and their trainers about group size and any implications for effective delivery or observed outcomes for pupils.

Limitations and lessons learned

Although originally conceptualised as a school-randomised trial, the design of the study was modified prior to EEF grants committee approval to a within-school, pupil-level randomised trial due to the practical considerations outlined in the introduction. While a pupil-level randomised trial solved the problem of adequately powering the trial for FSM pupils,

[53] 'Moderate evidence' indicates that a substantial body of evidence exists to support the effectiveness of small group tutoring but there may be limitations in the quality or quantity of some studies. This means that multiple rigorous studies have been conducted but may not be conclusive in all contexts.

this design placed artificial constraints on the selection of pupils and the dynamics of pupils in the intervention group. Since 1stClass@Number 1 is delivered to groups of four pupils within each school, schools were asked to nominate 12 to 16 pupils who were then screened using the SENT-R to select the final eight pupils for the trial. While the average SENT-R score of pupils selected for the trial was 31, the scores ranged from 0 to 63. This meant that, at least in some cases, schools nominated pupils who were less suitable for the intervention in order to maintain fidelity to the trial design (that is, the trial's requirement of a minimum of eight pupils from each school). These included pupils with severe mathematical difficulties who needed more intensive support than is provided by 1stClass@Number 1 and pupils with SENT-R scores greater than 49, who were less likely to benefit from the intervention.

Another consequence of the inclusion of pupils with very low or high SENT-R scores was the presence of pupils with vastly varying abilities within some intervention groups. Our IPE identified this as a challenge which made delivery difficult for some TAs as the content needed to be presented at different levels to different pupils. In usual implementation, schools might have chosen pupils of roughly similar abilities to receive the intervention in a group. A lesson from this exercise is that rather than use a SENT-R cut-off score to select pupils, it would have been more appropriate to select pupils from within a range of scores that had a lower bound commensurate with being able to access the support, for example. This would have helped to exclude pupils with very low prior attainment.

Furthermore, since the allocation of pupils to the intervention (and control) group was random and schools had been instructed to maintain this random allocation throughout the trial, schools did not have the opportunity to select pupils who would work well together as they would have done in usual implementation conditions. The pupil selection guidance provided by 1stClass@Number 1 to schools (when it is delivered in the usual context, outside of the trial) explicitly asks teachers to select pupils who 'gel' together in order to minimise time spent in managing the pupils and maximise time spent on the intervention. Our IPE notes the challenges experienced by a minority of TAs in managing pupil behaviour and group dynamics. However, overall, most Link Teachers felt that the pupil randomised design was easy to manage and reported that they would be willing to take part in another pupil-randomised trial.

One of our concerns with the pupil-randomised design was contamination, wherein control pupils were exposed to the content and pedagogical approach of 1stClass@Number 1. To mitigate this risk, we provided all schools with printable 'dos and don'ts' instructions for TAs. In addition to more obvious sources of contamination such as switching control and intervention pupils, the guidance also alerted TAs to slightly less obvious sources of contamination such as exposure of control pupils to displays of intervention pupils' work. As described under Usual Practice, less than ten of the 778 control pupils for whom maths intervention records were available received the full 1stClass@Number 1 intervention therefore concerns about widespread contamination were not confirmed or observed. Although the IPE indicated that there were a few instances where control pupils might have been exposed to displays of intervention pupils' work, in the absence of the rest of the intervention this is not likely to be a cause for concern.

As this was an effectiveness trial to assess the impact of 1stClass@Number 1 in real-world conditions, schools could choose to provide additional support to control pupils as per usual. A limitation of a 'business as usual' control condition (rather than an active control group) is that it does not allow us to disentangle the impact of the 1stClass@Number 1 intervention itself from the impact of pupils spending additional time receiving maths instruction in an intensive small group setting. An active control group—where pupils spent an equivalent amount of additional time on maths as 1stClass@Number 1 or were assigned to small group work in maths—would have helped to more directly address this problem. In attempting to power this trial for FSM pupils, an additional strategy we adopted was to oversample FSM pupils. Nationally, 22% of Year 2 pupils in state-funded schools in England are FSM-eligible. To ensure that we were adequately powering the trial for FSM pupils while also keeping the number of schools to be recruited at an achievable level, we assumed that, on average, 50% of pupils in the trial would be FSM-eligible; following our selection process, this figure was approximately 65%. The risk of oversampling FSM pupils is that the effect size for all pupils could be influenced by the effect size for FSM pupils. Furthermore, the higher proportion of FSM pupils in the trial sample likely does not reflect the proportion of FSM pupils among Year 2 pupils who would receive 1stClass@Number 1 in usual implementation conditions. This suggests that the findings of this trial may not be generalisable among the intervention's target population. However, as explained in the Findings section, our results show that the effect size for FSM pupils (0.11) is similar to that for all pupils (0.12) and indeed for non-FSM pupils (0.14), providing some reassurance that the primary analysis results of the trial are externally valid.

Key Stage 1 maths scores were used as the secondary outcome measure in the previous effectiveness trial of 1stClass@Number 1. However, the design of this trial allowed schools (at least in Cohort 1) to deliver the intervention to

control pupils after the completion of endpoint testing. This made it unfeasible to use any administrative data to measure broader impacts of the intervention on the maths attainment of pupils. Even if schools did not deliver the intervention to control pupils after endpoint testing, from the 2023/2024 academic year, the end of KS1 assessments were no longer mandatory, which might have meant a significant amount of missing data for the analysis.

Future research and publications

During the COVID-19 pandemic, EHU developed a blended training model, which included in-person and online training sessions for TAs. It has also developed a fully online training programme, which was delivered to some TAs in this trial. A blended or fully online training model would allow the delivery of 1stClass@Number 1 in schools that are located in regions where it may not be feasible for EHU ECC trainers to offer in-person training. Future research could involve robustly testing whether the intervention still works with the blended or fully online training model.

EHU plans to undertake further research examining the professional identities of TAs in primary schools. Specifically, it will explore how the experience of being trained and delivering a structured intervention like 1stClass@Number 1 influences TAs' professional identities. This body of work will involve interviews of TAs by EHU researchers and will also draw on the survey and case study data presented in this report.

References

Angrist, J. D. and Imbens, G. W. (1995) 'Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity', *Journal of the American Statistical Association*, 90 (430), pp. 431–442. Available at: https://doi.org/10.1080/01621459.1995.10476535

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015) 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, 67 (1), pp. 1–48. Available at: https://doi.org/10.18637/jss.v067.i01

van Buuren, S. and Groothuis-Oudshoorn, K. (2011) 'Mice: Multivariate Imputation by Chained Equations in R', *Journal of Statistical Software*, 45 (3), pp. 1–67. Available at: https://doi.org/10.18637/jss.v045.i03

Cole, S. R., Chu, H. and Greenland, S. (2013) 'Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer', *American Journal of Epidemiology*, 179 (2), pp. 252–260. Available at: https://doi.org/10.1093/aje/kwt245

Dong, N. and Maynard, R. (2013) 'PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies', *Journal of Research on Educational Effectiveness*, 6 (1), pp. 24–67. Available at: https://doi.org/10.1080/19345747.2012.673143

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007) 'School Readiness and Later Achievement', *Developmental Psychology*, 43 (6), pp. 1428–446. Available at: https://doi.org/10.1037/0012-1649.43.6.1428

EEF (2022) 'Statistical Analysis Guidance for EEF Evaluations'. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1738001240 (Accessed: 20 February 2025).

Farquharson, C., McNally, S. and Tahir, I. (2022) 'Education Inequalities. Available at: https://ifs.org.uk/sites/default/files/2022-09/Education-inequalities%20%281%29.pdf (Accessed: 19 December 2024).

Feaster, D. J., Mikulich-Gilbertson, S. and Brincks, A. M. (2011) 'Modeling Site Effects in the Design and Analysis of Multi-Site Trials', *American Journal of Drug and Alcohol Abuse*, 37 (5), pp. 383–391. Available at: https://doi.org/10.3109/00952990.2011.600386

Fox, J., Kleiber, C., Zeileis, A. and Kuschnig, N. (2021) 'ivreg: Instrumental-Variables Regression by "2SLS", "2SM", or "2SMM", with Diagnostics'. Available at: https://cran.r-project.org/web/packages/ivreg/index.html (Accessed: 29 July 2024).

Gerardi, K., Goette, L. and Meier, S. (2013) 'Numerical Ability Predicts Mortgage Default', *Proceedings of the National Academy of Sciences*, 110 (28), pp. 11267–11271. Available at: https://doi.org/10.1073/pnas.1220568110

Gerber S, A. and Green, P. D. (2012) *Field experiments: Design, Analysis and Interpretation*, New York: United States: W. W. Norton & Company. Available at: https://wwnorton.com/books/9780393979954 (Accessed: 23 January 2025).

GOV.UK (2023) 'Key Stage 1 and Phonics Screening Check Attainment, Academic Year 2022/23'. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-1-and-phonics-screening-check-attainment/2022-23 (Accessed: 19 December 2024).

Harrison, W. and Higgins, S. (2023) 'A Global Evidence Review for Policymakers: Findings from Meta-Analysis of Tutoring'. Available at: https://files.eric.ed.gov/fulltext/ED641960.pdf (Accessed: 19 December 2024).

Hedges, L. V. and Rhoads, C. (2010) 'Power Analysis in Education Research'. Available at: https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf (Accessed: 4 March 2024).

Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J. (2022) 'Valid t-Ratio Inference for IV', *American Economic Review*, 112 (10), pp. 3260–290. Available at: https://doi.org/10.1257/aer.20211063

Lord, P., Morton, C. and Sahasranaman, A. (n.d) 'Effectiveness Trial of 1st Class@Number 1: Statistical Analysis Plan'. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/1stClass%40Number1_Statistical-Analysis-Plan.pdf?v=1733086648 (Accessed: 20 December 2024).

Lord, P., Sahasranaman, A., Moore, E. and Morton, C. (2023) Effectiveness Trial of 1stClass@Number 1: Evaluation Protocol. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/EECN_Study_Protocol_OnlineTrainingUpdate_V2.pdf?v=1733086648 (Accessed: 20 December 2024).

Menzies, L., Ramaiah, B. and Boulton, C. (2021) A Space for Maths: Exploring the Need for Maths Tutoring and the Potential Role of Third Space Learning. Available at: https://cfey.org/wp-content/uploads/2021/09/A-Space-for-Maths-Sept-2021-for-publication-amended.pdf (Accessed: 19 December 2024).

Nunes, T., Barros, R., Evangelou, M., Strand, S., Mathers, S. and Sanders-Ellis, D. (2018) 1stClass@Number: Evaluation Report and Executive Summary. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/1stClass@Number_evaluation_report.pdf?v=1630925665 (Accessed: 4 March 2024).

Nunes, T., Bryant, P., Evans, D. and Barros, R. (2015) 'Assessing Quantitative Reasoning in Young Children', *Mathematical Thinking and Learning*, 17(2–3), pp. 178–196. Available at: https://doi.org/10.1080/10986065.2015.1016815.

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A. and Carraher, J. (2007) 'The Contribution of Logical Reasoning to the Learning of Mathematics in Primary School', *British Journal of Developmental Psychology*, 25(1), pp. 147–166. Available at: https://doi.org/10.1348/026151006X153127.

Parsons, S. and Bynner, J. (2005) Does Numeracy Matter More? Available at: https://dera.ioe.ac.uk/id/eprint/22317/1/doc_2979.pdf (Accessed: 19 December 2024).

Ritchie, S.J. and Bates, T.C. (2013) 'Enduring Links from Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status', *Psychological Science*, 24(7), pp. 1301–1308. Available at: https://doi.org/10.1177/0956797612466268.

Sharples, J., Webster, R. and Blatchford, P. (2021) Making Best Use of Teaching Assistants. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/teaching-assistants/TA_Guidance_Report_MakingBestUseOfTeachingAssistants-Printable_2021-11-02-162019_wsqd.pdf?v=1730378244 (Accessed: 19 December 2024).

Singh, A., Uwimpuhwe, G., Li, M., Einbeck, J., Higgins, S. and Kasim, A. (2022) 'Multisite Educational Trials: Estimating the Effect Size and Its Confidence Intervals', *International Journal of Research and Method in Education*, 45(1), pp. 18–38. Available at: https://doi.org/10.1080/1743727X.2021.1882416.

The R Foundation (2023a) R: the R project for statistical computing. Available at: https://www.r-project.org/ (Accessed: 4 March 2024).

The R Foundation (2023b) The R Project for statistical computing. Available at: https://www.r-project.org/ (Accessed: 15 August 2023).

Uwimpuhwe, G., Zhang, Q., Singh, A., Vallis, D., Einbeck, J., Higgins, S., Xiao, Z., De Troyer, E. and Kasim, A. (2020) 'eefAnalytics: Robust analytical methods for evaluating educational interventions using Randomised Controlled Trials designs, R software package.' Available at: https://durham-repository.worktribe.com/output/1699184/eefanalytics-robust-analytical-methods-for-evaluating-educational-interventions-using-randomised-controlled-trials-designs-r-software-package (Accessed: 29 July 2024).

Zeileis, A. (2006) 'Object-Oriented Computation of Sandwich Estimators', *Journal of Statistical Software*, 16(9), pp. 1–16. Available at: https://doi.org/10.18637/jss.v016.i09.

Zeileis, A., Köll, S. and Graham, N. (2020) 'Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R', *Journal of Statistical Software*, 95, pp. 1–36. Available at: https://doi.org/10.18637/jss.v095.i01.

Appendix A: EEF cost rating

Figure 2: Cost Rating

Cost rating	Description
£ £ £ £ £	*Very low*: less than £80 per pupil per year.
£ £ £ £ £	*Low*: up to about £200 per pupil per year.
£ £ £ £ £	*Moderate*: up to about £700 per pupil per year.
£ £ £ £ £	*High*: up to £1,200 per pupil per year.
£ £ £ £ £	*Very high*: over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity [0]		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low/Moderate	Confounding factors addressed by design. However, unpicking the intervention from 'additional maths time' was not done.
Threat 2: Concurrent Interventions	Low	Some suggestion of concurrent interventions for control pupils in some schools but not for intervention pupils.
Threat 3: Experimental effects	Low	Contamination risk due to pupil-level randomisation. Evaluators explore this and find some evidence of (positive) contamination but reported as very limited. Nothing reported on negative spillover – so assume not a problem. IPE confirmed that the guidance provided to schools on treatment allocation had been followed.
Threat 4: Implementation fidelity	Low	Programme implemented with high fidelity and with a relatively high rate of response to IPE data collection activities.
Threat 5: Missing Data	Low	13,2% missing data, MDES when accounting for it differs from primary estimate. However, evaluators provide a detailed missing data analysis which suggests no threat to validity.
Threat 6: Measurement of Outcomes	Low	Externally validated commercial measure
Threat 7: Selective reporting	Low	No evidence of selective reported. All the analyses are conducted in alignment with protocol and SAP. When adjustments to SAP have been made these have agreed with EEF and reported here.

- **Initial padlock score:** 4 Padlocks – within-school RCT
- **Reason for adjustment for threats to validity**: No adjustment made
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks

Appendix C: Changes since the previous evaluation[54]

Appendix Table 1: Changes since the previous evaluation[55]

Feature		Previous effectiveness trial	Current effectiveness trial
Intervention	Intervention content	One of the components of the intervention helps schools enhance or develop their approach to parental engagement in their child's mathematical learning. Children were given a mathematics activity called 'Special Delivery' every fortnight to take home and practice what they had learned. Parents were asked to help children do the activity and talk to them about it. A Post Office theme was used throughout the programme.	The formative evaluation by the BIT suggested that the Post Office theme could be less beneficial for disadvantaged pupils and recommended updating it to a new context that does not disadvantage FSM pupils. EHU updated the content of the Intervention to reflect a Parcel Delivery theme. 'Special Delivery' activities no longer refer to the Post Office.
	Delivery model	No change	
	Intervention duration	No change	
Evaluation	Eligibility criteria	1st Class@Number 1 is for Year 2 pupils who have fallen behind their peers and need more support at the level of the Year 1 curriculum. They may typically:	One of the main areas of focus in this trial was the pupil selection mechanism. In order to identify pupils eligible to participate in this trial, we followed a two-step process. In step 1, schools were asked to nominate 12 – 16 pupils (if possible) that they felt were eligible to receive 1CN1 (note, 10 was the minimum they could nominate, that is 8 plus a further 2 to mitigate against any attrition). In step 2, these nominated pupils were objectively screened using the Sandwell Early Numeracy Test (Revised) (SENT-R) to identify the final eight pupils selected to participate in this trial. The eligibility criteria for pupils to be nominated to the trial were similar to the previous trial. The changes are underlined. Suitable pupils may typically:

[54] Please delete this section if it is not applicable.

[55] Delete columns from the table if they are not applicable or adjust titles as relevant.

		• be able to count forwards and backwards in ones from 10 • have some knowledge of number facts • be able to perform simple addition and subtraction calculations using counting all and counting out approaches • be able to read and write numbers but lack secure understanding of their magnitude and quantity • struggle with mental calculation strategies and mathematical vocabulary • lack confidence in mathematics and be reluctant to talk about their mathematical learning. Teachers were also advised to nominate for the same group children who can 'gel' and learn together and were not participating in another intervention, in order to avoid overload. The intervention team gave guidance to schools on how they might take into account the pupils' scores in the QRT pre-test as part of the information to be considered when they nominated	• be able to count forwards in ones to 10 • have some knowledge of number facts and also have some understanding of the composition of number within 10 • be able to perform simple addition and subtraction calculations using counting all and counting out approaches • be able to read and write numbers but lack secure understanding of their magnitude and quantity • struggle with mathematical vocabulary • lack confidence in mathematics and be reluctant to talk about their mathematical learning. Additionally, because the trial has a focus on disadvantaged pupils, schools were encouraged to nominate FSM-eligible pupils who met the above criteria. To avoid overload, pupils participating in other maths interventions were not eligible for this trial. Because of the pupil-level randomisation, schools were not specifically asked to select pupils who could 'gel' and learn together. In the next step, the SENT-R was administered 1:1 to all nominated pupils. NFER shared a customised report with each school rank ordering pupils based on their raw SENT-R scores and recommending eight pupils for selection to the trial. Schools were asked to confirm the final eight participants to be selected to the trial. In recommending pupils for selection to the trial, we first prioritised the selection of FSM-eligible pupils with raw SENT-R scores of 49 or below. While pupils scoring 40 or below were considered to be around 12 months behind age-related expectations, in agreement with EHU we also allowed the selection of pupils with a raw score between 41 – 49, where a raw score of 49 equates to a number age of 6:10 (this is part of usual practice that EHU experience). Once FSM-eligible pupils were recommended, we randomly chose non-FSM pupils with raw scores of 40 and below to reach the target of eight pupils for recommendation to the trial. If this target was not achieved by choosing pupils with scores of 40 or below, we then randomly chose non-FSM pupils with a raw score of 41 - 49 in order to reach the required number of eight pupils. In the event that a school did not have eight pupils who fell below the score of 49 on the SENT-R, we randomly chose pupils with scores above 49 to reach the target of eight pupils (this was preferable to randomising fewer than eight pupils in a school, further details below). Schools were asked to briefly confirm the recommended shortlist and also given the option of replacing recommended non-FSM pupils with other pupils who were not recommended by us.

		pupils for participation in the project.	
	Level of randomisation	School-level randomisation	Pupil-level randomisation within schools
	Outcomes and baseline	No change in baseline and primary outcome. The secondary outcome was Key Stage 1 (KS1) maths. KS1 maths results were obtained through the NPD using the variable KS1_MATH_OUTCOME.	We were unable to use KS1 maths as a secondary outcome in this study for two reasons. In summer 2024, this test was no longer mandatory in schools. Furthermore, the trial schools may have delivered 1st Class@Number 1 to control pupils following the trial period, but before any longer-term attainment outcome measurement would be possible.
	Control condition	Control condition remains business-as-usual	Although the control condition remained the same, because this is a pupil-randomised trial, there is a risk of contamination that was not present in the previous trial. However, this risk was deemed low, given that the TA delivered the intervention to pupils in a small group outside of the classroom. TAs has a register for the pupils in the intervention group. Control pupils remained in their normal classroom setting. Schools were specifically asked not to switch pupils' allocation. It was therefore unlikely that control group pupils received any aspects of the intervention.

Appendix D: Effect size estimation

Appendix Table 2: Effect size parameters from the primary analysis and FSM subgroup models

		Variance components obtained from a model with no predictors		
Population	Adjusted differences in means	Between-school variance σ_B^2	Within-school variance σ_W^2	Effect size denominator $\sqrt{\sigma_B^2 + \sigma_W^2}$
All pupils	0.52	6.86	12.52	4.40
FSM pupils	0.48	6.46	12.43	4.35

Appendix Table 3: Effect size parameters from the sensitivity analysis that includes a random effect of the intervention

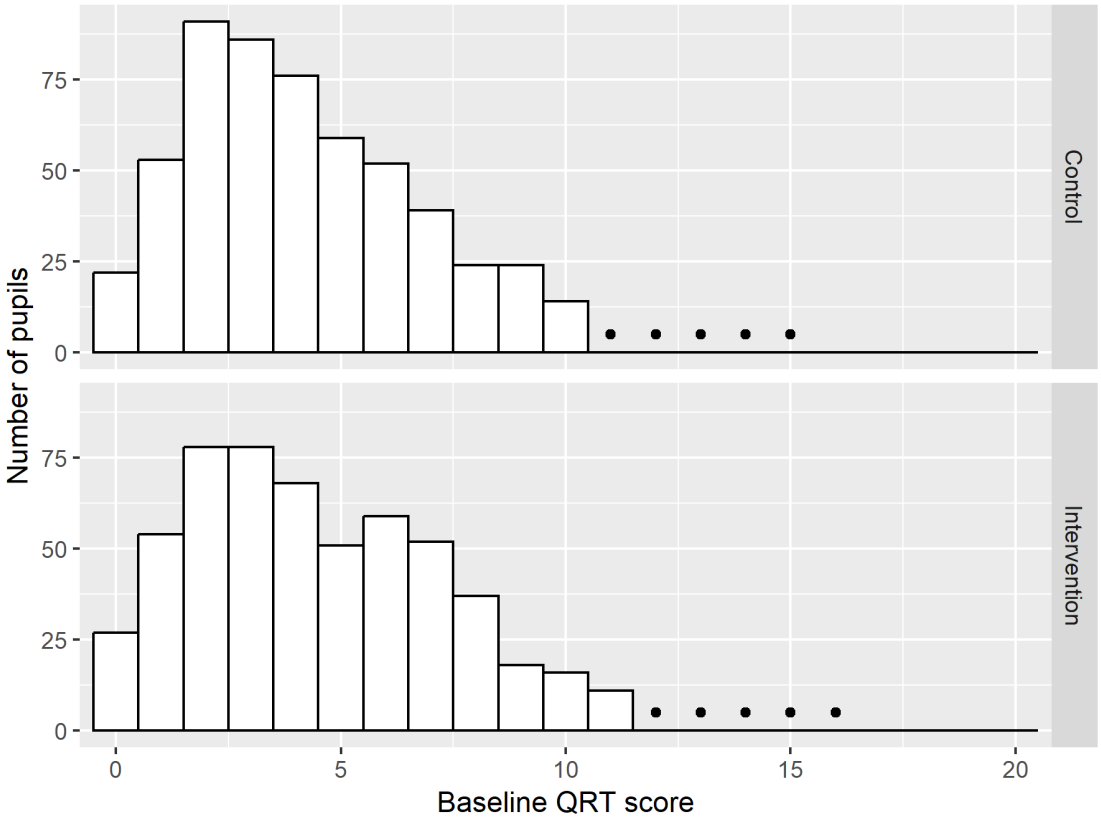
		Variance components obtained from a model with no predictors		
Population	Adjusted differences in means	Between-school variance σ_B^2	Within-school variance σ_W^2	Variance due to differing intervention effect between schools σ_E^2
All pupils	0.52	7.04	11.96	1.97
FSM pupils	0.48	7.55	12.04	1.29

Appendix E: Survey items measuring TA's confidence is their own maths abilities

D	YOUR CONFIDENCE IN YOUR MATHS ABILITIES					
Please select one response for each statement.		1 Strongly disagree	2 Disagree	3 Agree	4 Strongly agree	5 I don't want to say
D.1	My mental maths abilities are good					
D.2	I am confident solving maths problems					
D.3	I am confident using maths in my daily life (for example estimating a bill; calculating a 20% off deal)					
D.4	I believe I'll be able to understand the content delivered in maths CPD sessions					
D.5	I believe I am good at maths					
D.6	Overall, I feel confident in my own maths abilities					

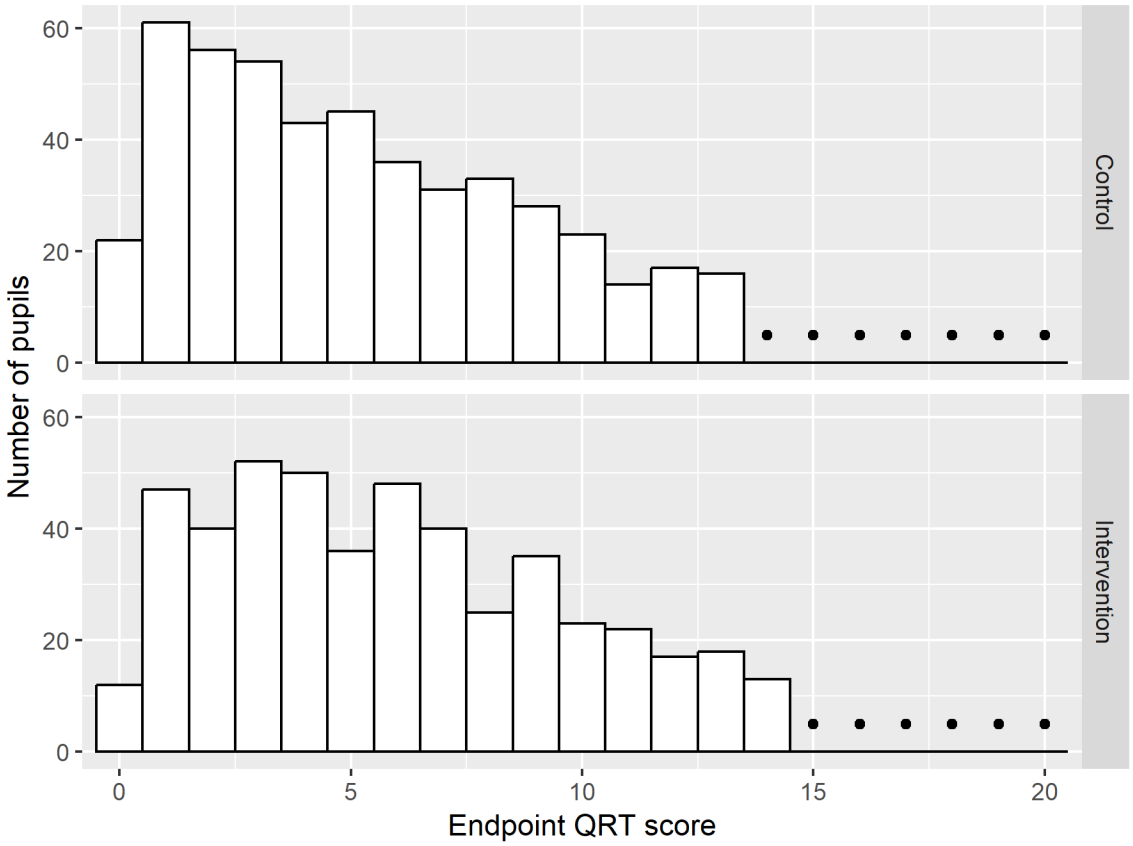
Appendix F: Distribution of QRT scores in the subgroup of FSM-eligible pupils

Appendix Figure 1: Distribution of baseline QRT scores amongst FSM-eligible pupils in the control and intervention groups



Black dots on the histogram represent pupil counts greater than zero but less than ten. These have been supressed to ensure individual pupils cannot be identified from the figures in this report.

Appendix Figure 2: Distribution of endpoint QRT scores amongst FSM-eligible pupils in the control and intervention groups



Black dots on the histogram represent pupil counts greater than zero but less than ten. These have been supressed to ensure individual pupils cannot be identified from the figures in this report.

Appendix G: Results from the missing data logistic regression models

Appendix Table 426: Results from two multilevel logistic regression models in which the outcome is whether a pupil's baseline QRT score was missing (1) or not missing (0)

Population	Variable	Estimate	SE	Odds ratio (95% CI)	p-value
All pupils	Group=Intervention	-0.38	0.44	0.69 (0.29, 1.64)	0.396
	Endpoint QRT score (per point)	-0.11	0.08	0.9 (0.77, 1.05)	0.173
	FSM=Yes	0.21	0.54	1.23 (0.43, 3.55)	0.702
	SEN=Yes	0.17	0.51	1.18 (0.43, 3.23)	0.747
	SENT-R score (per point)	-0.06	0.03	0.94 (0.89, 1)	0.066
	Establishment type=Free Schools	1.97	2.70	7.21 (Not estimable)	0.465
	Establishment type=Local authority maintained schools	0.67	0.86	1.95 (0.36, 10.54)	0.440
	FSM quintile=Lowest 20%	0.32	2.34	1.37 (Not estimable)	0.892
	FSM quintile=2nd Lowest 20%	0.03	1.43	1.03 (0.06, 16.86)	0.984
	FSM quintile=Middle 20%	-2.34	2.60	0.1 (0, 15.85)	0.369
	FSM quintile=Highest 20%	-0.07	0.90	0.93 (0.16, 5.43)	0.936
FSM pupils	Group=Intervention	-0.88	0.65	0.41 (0.12, 1.48)	0.176
	Endpoint QRT score (per point)	-0.11	0.14	0.9 (0.69, 1.17)	0.416
	SEN=Yes	-0.79	0.87	0.45 (0.08, 2.49)	0.364
	SENT-R score (per point)	-0.07	0.06	0.93 (0.84, 1.04)	0.212
	Establishment type=Free Schools	4.26	5.29	Not estimable	0.421
	Establishment type=Local authority maintained schools	1.46	3.15	4.29 (Not estimable)	0.643
	FSM quintile=Lowest 20%	Not estimable	Not estimable	Not estimable	1.000
	FSM quintile=2nd Lowest 20%	1.60	3.90	4.97 (Not estimable)	0.681
	FSM quintile=Middle 20%	Not estimable	Not estimable	Not estimable	1.000
	FSM quintile=Highest 20%	0.64	2.84	1.89 (Not estimable)	0.823

Appendix Table 527: Results from two multilevel logistic regression models in which the outcome is whether a pupil's endpoint QRT score was missing (1) or not missing (0)

Population	Variable	Estimate	SE	Odds ratio (95% CI)	p-value
All pupils	Group=Intervention	0.15	0.16	1.16 (0.85, 1.58)	0.349
	Endpoint QRT score (per point)	0.01	0.03	1.01 (0.96, 1.07)	0.638
	FSM=Yes	0.32	0.19	1.38 (0.96, 1.99)	0.085
	SEN=Yes	0.08	0.17	1.08 (0.77, 1.52)	0.650
	SENT-R score (per point)	-0.01	0.01	0.99 (0.97, 1.01)	0.405
	Establishment type=Free Schools	-0.44	0.87	0.64 (0.12, 3.53)	0.611
	Establishment type=Local authority maintained schools	-0.36	0.19	0.7 (0.48, 1.01)	0.056
	FSM quintile=Lowest 20%	0.10	0.57	1.1 (0.36, 3.37)	0.865
	FSM quintile=2nd Lowest 20%	-0.16	0.36	0.86 (0.42, 1.74)	0.665
	FSM quintile=Middle 20%	-0.22	0.32	0.8 (0.43, 1.49)	0.484
	FSM quintile=Highest 20%	-0.17	0.22	0.84 (0.55, 1.29)	0.426
FSM pupils	Group=Intervention	0.16	0.19	1.18 (0.81, 1.72)	0.398
	Endpoint QRT score (per point)	-0.01	0.04	0.99 (0.92, 1.07)	0.892
	SEN=Yes	0.06	0.21	1.06 (0.7, 1.6)	0.778
	SENT-R score (per point)	0.01	0.01	1.01 (0.98, 1.03)	0.674
	Establishment type=Free Schools	-0.66	1.16	0.52 (0.05, 5.02)	0.570
	Establishment type=Local authority maintained schools	-0.28	0.21	0.76 (0.5, 1.15)	0.193
	FSM quintile=Lowest 20%	Not estimable	Not estimable	Not estimable	0.997
	FSM quintile=2nd Lowest 20%	-0.17	0.48	0.84 (0.33, 2.17)	0.722
	FSM quintile=Middle 20%	-0.23	0.39	0.79 (0.37, 1.69)	0.547
	FSM quintile=Highest 20%	-0.22	0.24	0.8 (0.5, 1.27)	0.348

Further appendices

Please see accompanying document *'Further Appendices'*.

This document is available for download at https://educationendowmentfoundation.org.uk

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London SW1P 4QP

@EducEndowFoundn

Facebook.com/EducEndowFoun